Ancient Temple Pillar Segmentation Using a Fully Convolutional Neural Network Model

Gurudeva Shastri Hiremath*[1], Shrinivasa Naik C. L. [2], Narendra Kumar S. [3]

Abstract: The historical temples in India belonged to illustrious kingdoms that ruled for nearly a millennium. Their pillar style, sculpture, inbuilt, architecture, technique, vastness and magnitude have very awesome wonders of their own. Using useful information from the on-site diagnostic of the raw history of the pillar architecture, archaeologists can make decisions about many aspects of pillar handling and management techniques. Archaeologists can better understand old temples by segmenting the pillars, which is useful for future research like identification & classification of different types of pillars based on architecture, to know the original architecture adopted during the construction of temples by various dynasties, and this original architecture information guides the re-construction of temples. Because there are no reliable digital methods for automatic pillar segmentation, archaeologists must deal with a number of challenging issues. Due to irregularities in image acquisition, complex architectural designs, noise, time, and imaging distortions, automated pillar segmentation presents difficulties. In the literature, certain inaccurate statistical segmentation techniques for pillar segmentation have been suggested. For the auto-segmentation of pillars, we suggest a fully convolutional network(FCN) Model in this paper. The suggested technique reduces the unpredictability of picture noise and develops FCN models using images from our own generated dataset. Furthermore, optimal data augmentation and model hyperparametrization are shown to prevent overfitting for pillar area segmentation. With a recall/precision rate of 0.9698/0.9200, the proposed approach is examined on the test dataset. When compared to published algorithms in the literature segmentation challenge, the new method performs better, with a Dice correlation coefficient of 0.9284, than those algorithms.

Keywords: Pillar Architecture, Auto-Segmentation, Convolutional Neural Networks, Deep Learning.

1. Introduction

According to Kenyon in 1940, archaeology is "The science of the treatment of the material remains of the human past." [1]. The various Archaeologists study human past history and first development of stone. It is particularly about prehistoric monuments. Archaeology may show how a civilisation developed, its environment, culture, religion, and much more[2]. From the 19th century to the present, archaeologists have studied the history of the Indian subcontinent, which encompasses a wide range of archaeologists. Western European visitors were the first scholars to become interested in the archaeology of the Indian subcontinent in the early 18th century [3]. Alexander Cunningham, the first director of the Archaeological Survey of India, had a significant impact on academic research into Indian archaeology. India has rich information of ancient monuments of the archaeological and the historical importance. There are so many ancient monuments are built by number of empires or ruling dynasties time even after nearly thousands of

years. Their style, sculpture, inbuilt, vastness, architecture, magnitude and technique define their more unique way of structuring. The notable ruling dynasties of India included the Kadambas, Gangas, Chalukyas of Badami and Kalayana, Rashstrakutas, Hoysalas, and Sultanate empires [4]; they were the primary builders of the Ancient Indian Temples. The vimana (goupara), which is Sanskrit for "temple," and pillars are the two most significant structural components of a temple. The pillars (Stambha), though there was no software, calculator, or spreadsheet when these temples were constructed thousands of years ago, are a crucial component in Hindu temple architecture. The final products exhibited by these temples are fantastic and are "Structural Engineering Marvels." [5].

Currently Identification & classification of pillars are defined by the human visual abilities, whereas most of the foremost features are minute and difficult to understand. Referable to this complexity of architecture features, segmentation of pillar plays a very important role. Proper segmentation will address this complexity by fetching important features, those maps to particular architecture. Archaeologists face many complex problems in this process because do not have no such authentic auto-segmentation of pillars. Automatic diagnosis and scientific segmentation of pillars are highly desired by archaeology experts. Referable to the complexity of architecture, the

---

[1] St. Joseph Engineering College, Mangaluru-575028 , Visvesvaraya Technological University, Belagavi-590018, INDIA.
[2] U.B.D.T College of Engineering, Davanagere-577004, Visvesvaraya Technological University, Belagavi-590018, INDIA.
[3] J.N.N College of Engineering, Shivamoga-577204, Visvesvaraya Technological University, Belagavi-590018, INDIA.
* Corresponding Author Email: devguruphd4u@gmail.com

techniques of computer vision based auto-segmentation appears to be well adjusted.

In image processing, segmenting images is a key approach. Pattern recognition, computer vision, and image analysis all use it as a pre-processing stage [6]. Splitting a digital image into several portions (pixel sets) is the process of image segmentationImage segmentation is a method frequently used to recognize objects and boundaries (lines, curves, etc.) in photographs. Image segmentation results in a series of blocks that span the entire image or, in the case of edge detection, a series of contours extracted from the image. Every pixel in an area of an image is connected to a specific characteristic or calculated property, like color, intensity, or texture[7]. Neighbouring regions diverge substantially in terms of the same characteristics. Image segmentation has a variety of applications, including digital libraries, image processing, medical imaging, computer vision, face recognition, photo and video retrieval, etc.[8]. There are various methods for segmenting images, including thresholding, artificial neural networks, region-based, edge-based, and clustering-based segmentation, as well as feature-based segmentation. In this section, an analysis is done on how to extract region-of-interest artefacts or objects from the pre-processed image in order to solve the pertinent issue. The pre-processed image will develop noise reduction/removal, edge identification, and enhancement in accordance with our research effort. A survey on edge-based , threshold-based and CNN based segmentation is conducted in order to extract the crucial component of the pillar based on edge detection.

Rossella Cossu et.al provide colour picture segmentation mechanism based on edge detection and histogram threshold to separate degradation, which is represented by cavities or holes, from various coloured stone-material images[9]. The general recursive method for picture segmentation is described in by expanding Otsu's method. A fresh and unique procedure has been established with regard to the assortment of document photos, particularly real bank checks. By segmenting the brightest homogeneous object from a picture after each iteration, this approach leaves just the darkest homogeneous object after the last recursion. The same procedure could be applied to strengthen digitized essential elements by removing undesired parts[10]. Fernando C et.al proposed a new picture segmentation approach that emphasizes edge and region-based information using a spectrum method and the watershed morphological algorithm. An image's noise is first reduced using a bilateral filter as a pre-processing method, after which region merging is introduced to perform the initial segmentation, region similarity is created, and graph-based region grouping is completed using the Multi-class Normalised Cut method[11, 12]. Shiping Zhu et al. published a novel threshold-based

technique for edge detection and picture segmentation. They use the pixels in its immediate vicinity to determine each pixel's threshold. With the aid of the suggested technique, they also locate the image's edges. A threshold for each pixel was determined using a histogram. Use PDF to separate the threshold and background of the image. This method could be modified to extract the intricate information from images of historical monuments [13].

The segmentation process was made more effective with the use of several methods, including active contour, level set, fuzzy clustering, and k-means clustering. Author conducts a thorough analysis of picture segmentation algorithm performance. The level set approach has two levels: intensity-based image segmentation and texture-based image segmentation. Combining intensity and texture-based image segmentation with traditional approaches produces better results. [14]. The rapid marching numerical method is used to introduce the novel colour image segmentation technology. Only the contour of the area under study was used in the implementation of this technique by Maria Mercede Cerimele & Rossella Cossu. They also offered a more comprehensive idea, which was the ability to separate decay zones from the entire image; despite their geographic separation, these regions all had comparable colorimeter values [15]. A superior picture segmentation methodology based on a level sets method can be used to identify deterioration on a building's surface. This segmentation technique lowers the expense of monitoring while also subjectivizing the outcome [16]. Comparative analysis of the structural characteristics of temples using VOE and DSC metrics shows that regional segmentation is more accurate than the threshold method[17].

The results of using convolutional neural networks (CNNs) to segment biological images are promising when compared to those of older techniques[18,19]. CNNs are a unique type of neural network model that use repeated convolution operations to extract characteristics particular to the segmentation task from an input image. Convolution filters are provided by the neurons' learnable biases and weights. Architectural solutions for various segmentation issues can be designed using different arrangements of these filters. One such important CNN model that is utilized for a variety of various medical image segmentation tasks is the U-Net architecture presented by Ronneberger et al.[20]. The fully convolutional network (FCN) known as U-Net was made popular by Long and Shelhammer [21]. The output is the relevant segmentation mask if an input image is given. Milletari et al. [22] altered their model to include 3D photo stacks as input, further developing this concept. They also created the concept of residual connections, which is covered in the well-known ResNet architecture [23]. Our selection of FCN in manufacturer-independent pillar segmentation challenges

is driven by CNNs' capacity for abstract learning and self-learning to recognise minute spatial differences. The pillars of old temples have not yet been segmented using FCNs.

(a)   (b)   (c)

(d)   (e)
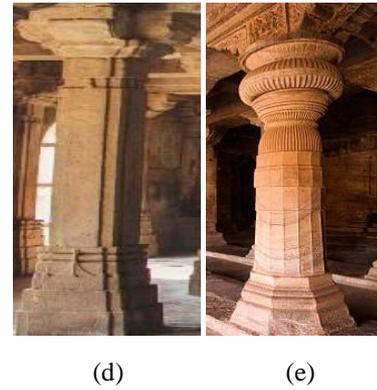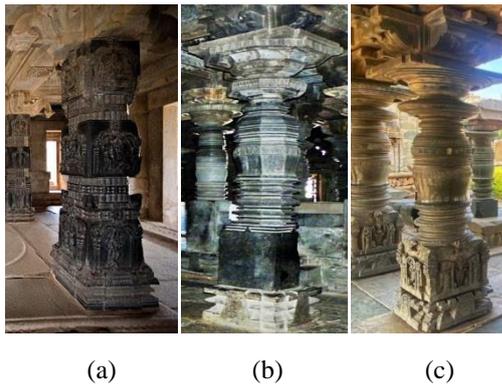
Fig. 1. Sample Images of Pillars (a)Bhramakantha, (b)Vishnukantha, (c)Rudrakantha, (d)Bhanukantha (e) Soumyakantha.
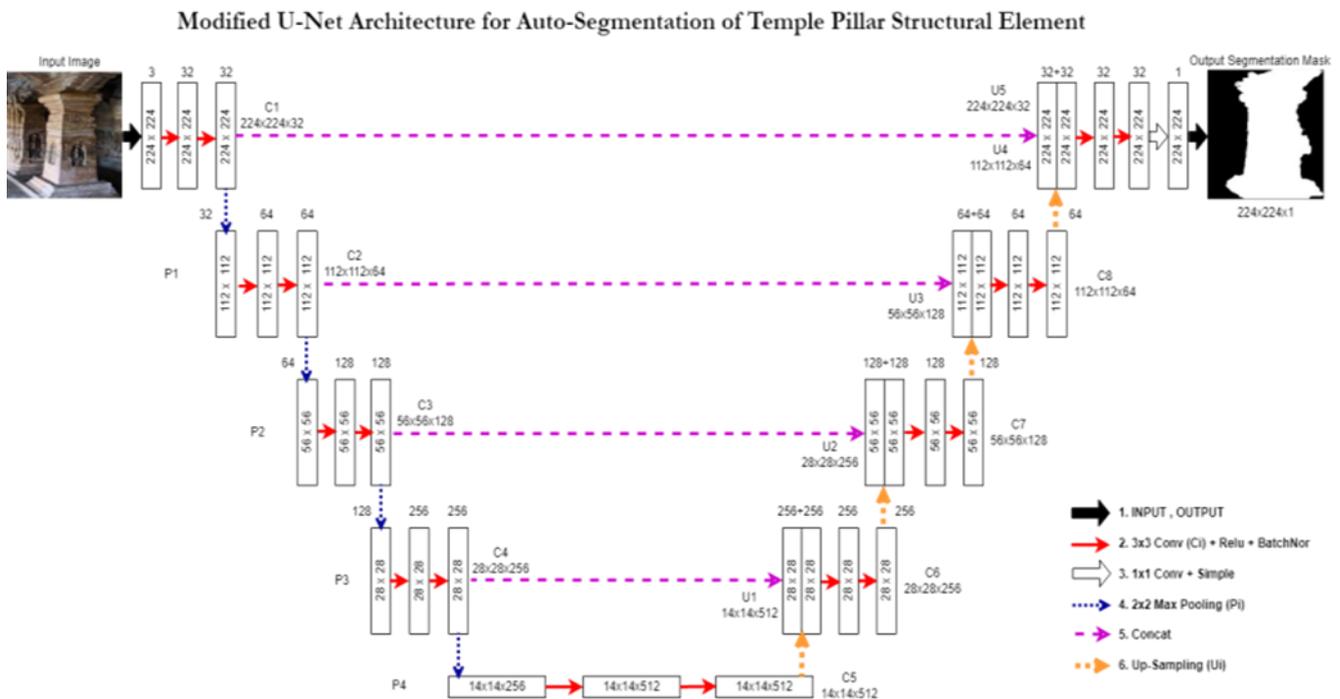
Fig. 2. The proposed network architecture with between 32 and 512 kernels make up the 18 convolutional layers of the FCN model.

To categorise the various types of temple pillars and the architectural styles employed in their construction, archaeologists can identify the structural characteristics of the pillars using automated segmentation methods. This will help them make better decisions while rebuilding the temple. We chose to use CNN-based segmentation approaches in this work since the literature review revealed that they outperformed the other segmentation methods in terms of performance. In the recent literature, a number of CNN segmentation techniques have been put forth. To the best of our knowledge, the acquisition and visualization-specific parameters are chosen using currently used methods, which depend on human intervention. The noise level and various pixel intensities in the pillar images change as images are captured at various times, making pillar segmentation a difficult operation. Fig. 1 displays

some images of pillars that were taken based on how severe the shapes and ornaments (ornaments) were. These issues that emerge in the creation of an automated independent pillar segmentation method are addressed by the proposed approach.

Two key contributions are made by this work. First, an adaptable, cutting-edge FCN model that can divide pillars automatically from images of pillars that also include undesired backgrounds is shown. We investigate the sensitivity of model parameters like core dimensions and layer count to the goals of pillar area segmentation. We discover that, when compared to cutting-edge methods, a properly parameterized model may achieve a higher recall rate of 0.9698 while maintaining the precision rate of 0.9284 across diverse pillar kinds.

The remaining portions of the work are organized as follows. The methodology and data are covered in Section II. In Section III, the experimental settings are explained. Section IV presents and discusses the findings of the experiment. Section V concludes with discussions.

2. Methods and Data

The proposed work is split into two phases: a pre-processing phase that includes image noise removal and a segmentation phase for pillars. Due to the noise that naturally accompanies pillar images, automated methods for segmenting pillars perform less well. Pillar segmentation performance is enhanced by the first preprocessing module, which eliminates this undesirable noise. The next two methodological steps are discussed.

2.1. Pre-processing: Denoising as Erosion and Dilation

Like other images, pillar images have different amounts of noise. Morphological procedures can be used to approximate the noisy images. We apply the erosion and dilation method proposed in [24] to denoise the pillar image data set in response to this study. A group of operations known as morphological operations transform images based on forms. By including a structural feature to an input image, you can produce an output image. Two morphological processes—erosion and dilation—are the most fundamental.

Basics of Erosion:

- Erodes away the boundaries of the foreground object
- Used to diminish the features of an image.

Basics of dilation:

- Increases the object area
- Used to accentuate features

The new morphological layer known as the PConv layer, which is based on the CHM filter formulation, is now ready to be introduced. The PConv layer executes the following procedure for an individual channel image $f(x)$ and individual filter $w(x)$:

$$PConv(f; w; P)(x) = \frac{(f^{P+1} * w)(x)}{f^P * w}$$
$$= (f * P w)(x) \qquad (1)$$

P, a scalar that determines the type of operation (P < 0 pseudo-erosion, P > 0 pseudo-dilation, and P = 0 standard linear convolution), and the weighting kernel $w(x)$, whose corresponding asymptotic structuring function is given by $w(x) = \log(w(x))$, are the parameters that define the system. Since these parameters are differentiable in this formulation, we may utilize gradient descent. Refer to "(1)".

2.2. Pillar Segmentation: FCN Model

Denoised pillar pictures are input into the FCN model [25], which produces a predictive value that is later used to produce a binary output mask matching to segmented parts. Inspiration for our FCN approach comes from Ronneberger's U-Net architecture[26], which builds a precise segmentation map using local and global data from an input picture. Local characteristics determine the specific pillarar borders while global features specify the precise location and comparative size of the pillarar region.

The proposed architecture employs a two-phase, multi-level strategy. Convolutions are performed in the first phase in five steps. To obtain a bigger receiving field, each stage uses two layers of convolution followed by a layer of downsampling with a max-pooling operation. The second stage seeks to return the first phase's activations to the initial solution. Utilising trainable unfolding layers, this upsampling is carried out in four stages and makes consistent predictions regarding the position of the pillars. Both local and global characteristics are taken into account for exact segmentation. As a result, in phase one's forward and concatenation steps, In phase two, the features from the upsampled stages' matching output are mixed with the characteristics that were extracted from the preceding stages. This architecture has the benefit of treating each pixel as a single training example with a single backpropagation mistake. This procedure greatly expands the training data set, preventing the model from becoming overfit.

In Fig. 2, the suggested network architecture is displayed. The model is made to handle images with a resolution of [256 x 512]. In the entire network, [3X3] is chosen as the convolution filter's width and height. Based on earlier research [27], this filter size ensures reduced parameterization compared to bigger filter sizes while ensuring learning of discriminating features from the pixel neighbourhood. In the analysis/downsampling stage, two [3X3] convolutions are placed in front of a [2X2] max-pooling layer, bringing the image's resolution exactly down to half. A batch normalisation operation and a ReLU (Rectified Linear Unit) activation function are applied after each convolution. Batch normalisation has the effect of accelerating network training and convergence [28]. The number of filters (k) in each convolutional layer also doubles after each step. The activations are upsampled using deconvolution or fractional strided convolution in the network's second phase. It travels a path symmetric to the downsampling strategy, except deconvolutions are used in place of the pooling operations. A [1 x 1] convolutional layer makes up the network's last layer. The binary cross-entropy loss is used to convert the input to a probability map with the same dimensions as the input image, and the output of the final [1 x 1] layer is utilized to build the

network's loss function. The loss function of the network for a single input image with no pixels is shown in (2).

$$L = -\sum_{i=1}^{n_{out}} (t_i \log(s_i) + (1 - t_i) \log(1 - s_i)) \qquad (2)$$

where t_i is the actual binary output (target) and s_i is the predicted binary output for pixel i. The sigmoid activation function S and the weighted sum of the inputs and final output, y, are shown in (3).

$$s_i = \frac{1}{1 + e^{-y_i}} , \qquad y_i = \sum_{j=1}^{n} x_j w_{ji} \qquad (3)$$

The letter n here represents the number of neurons in the layer just before to the sigmoid activation. As a result, the model's output shows the likelihood of each individual pixel being a pillar or not. The target t pillar mask is then compared to the generated binary image mask, and model weights are changed using the loss function L to guarantee the loss is kept to a minimum for succeeding epochs. The FCN model has received such training. The proposed model [see Fig. 2] features 18 convolutional layers in total throughout the two stages, which are represented by red arrows in the diagram. There are 32 filters (k) assigned to the initial two layers. The next two convolutional layers have twice as many filters after each pooling layer. A convolutional layer can have a maximum of 512 filters because there are 4 pooling layers. During the upsampling phase, the quantity of convolutional filters is reduced by half after each upsampling operation. This is repeated four times, just like in the previous phase, k = 32 filters will be used in each of the final two convolutional layers. The next step is to add a [1 X 1] convolutional layer with a filter to produce an output picture mask with [256 X 512] dimensions that match the input dimensions.

2.3. Data

On our own produced dataset of ancient temple pillars, to evaluate the effectiveness of the proposed technique, experiments are carried out. Based on the severity levels of the shapes and ornamentations, there are five different types of pillars in this collection of ancient temple pillars. This dataset has only been provided for research purposes via our JNNCE college web site and the Kaggle web portal [29, 30].

Table 1. Dataset Description

Sl.No	Pillar Type	All 5 Types of Pillar Images Spilt-up as			Total number of Images for each Type.
		Train set	Valid set	Test set	
1	Bhanukantha	1055	300	150	400
2	Bhramakantha	1055	300	150	231
3	Rudrakantha	1055	300	150	373
4	Soumyakantha	1055	300	150	173
5	Vishnukantha	1055	300	150	328
Total Number of Images in Ancient Temple Pillar Dataset.					**1505**

The identities of the various pillar styles have been verified by archaeologists. Each image in the newly constructed image dataset is linked to a professional label showing a variety of categories based on the intensity of the Shapes and Ornamentations, including Bhramakantha, Vishnukantha, Rudrakantha, Bhanukantha, and Soumyakantha as shown in Fig. 2 [5]. Training and test subsets of this dataset have been created. The dataset description is shown in Table 1. Fig 1 displays an example of a Pillar image from each category.

3. Experimental setup

Keras 1.0 incorporates the suggested technique DCNN models architecture [31]. The server running Ubuntu Linux includes a 3.40 GHz i7-3770 processor, 16 GB of memory, and an 8 GB GTX 1070 GPU.

### 3.1. FCN Model Training and Hyper-Parametrization

For the intended outcomes, model hyper-parameters must be optimally calibrated. The model's learning rate, the number of layers, the number of filters and/or kernels in each layer, the number of weights and biases, are the most important parameters/hyper-parameters that need to be optimized in the context of FCNs. By employing the hold-out approach in grid-search, the optimal combination of these parameters is found [32].

Table 2. Various FCN architectures made with a variety of layers.

Sl.No	No of Layers	Initial Filter Size ($\omega \times \eta \times k$)	Depth	No of Parameters	Time elapsed (hh:mm:ss)
1.	4	$3 \times 3 \times 64$	0	38,913	00:56:55
2.	6	$3 \times 3 \times 64$	1	408,321	00:54:25
3.	10	$3 \times 3 \times 64$	2	1,884,417	00:53:45
4.	14	$3 \times 3 \times 64$	3	7,785,729	00:57:55
5.	18	$3 \times 3 \times 64$	4	31,384,833	01:13:13

Finally, using the test data set, a performance analysis of the trained model is carried out. The FCN model architecture as described in Section 2.2 should be validated, a number of experiments have been developed. A max-pooling layer or a deconvolution layer should always be placed after two convolutional layers, according to [26]. In order to preserve this feature and the number of filters indicated in [26], the number of layers can change for different architectures. Table II provides a summary of various FCN architectures. Max-pooling layers are indicated by depth, weights and biases are indicated by parameters, and convolutional layers are shown by layers. Training sets, validation sets, and test sets each comprise 1055, 300, and 150 images from each of the five kinds in order to evaluate the FCN parameters experimentally on the different designs. Erosion and Dilation pre-processing activities are carried out before delivering images as inputs to the FCN network, as was covered in Section 2.1. The images are then scaled down to a [256X512] resolution standard using bilinear interpolation. The FCN network is trained using the binary cross-entropy loss function (see (2)). The stochastic gradient descent-based Adam optimizer is used to update the weights [17]. The default hyper-parameters for this optimizer are $\beta 1 = 0.9$, $\beta 2 = 0.999$, and $\epsilon = 1 \times 10-8$. A brief random search is used to empirically determine the learning rate, which is $3 \times 10-4$. The He initializer is used to determine the initial network weights [33]. The approach doesn't need any additional pre-trained weights. The FCN network is completely trained from scratch on the train data set. Nevertheless, there are not enough examples in the train data to fully train an FCN. Thus, data augmentation is employed to increase the applicability of the model and decrease over-fitting. On training loss and model convergence, the effects of several data augmentation approaches are examined. These include fluctuations in brightness and contrast as well as edge and gradient enhancement for images.
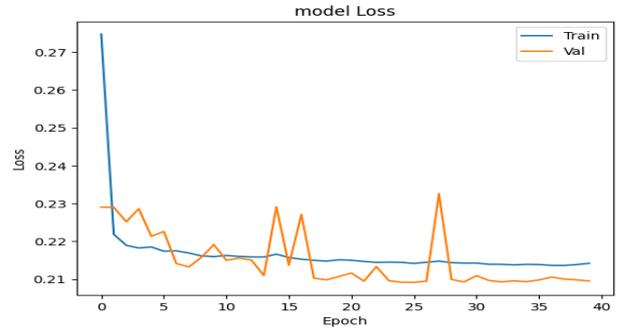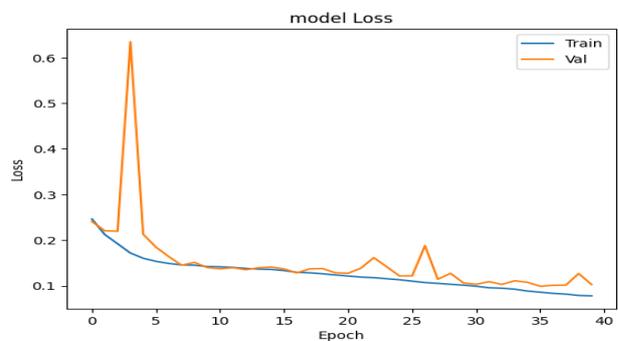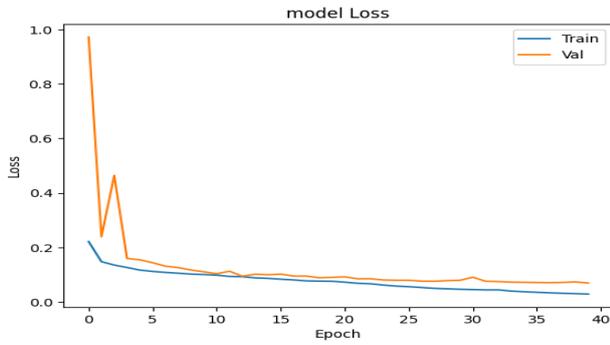

(a) Layers = 4, Depth = 0.
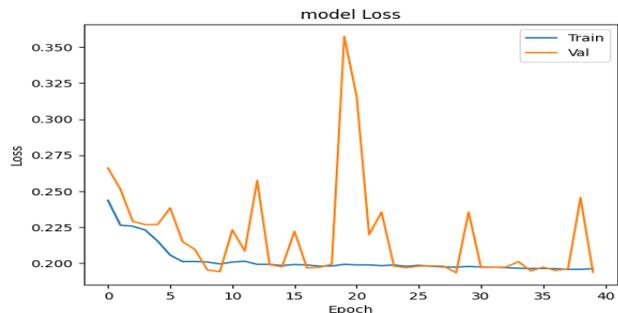
(b) Depth = 1, Layers = 6.
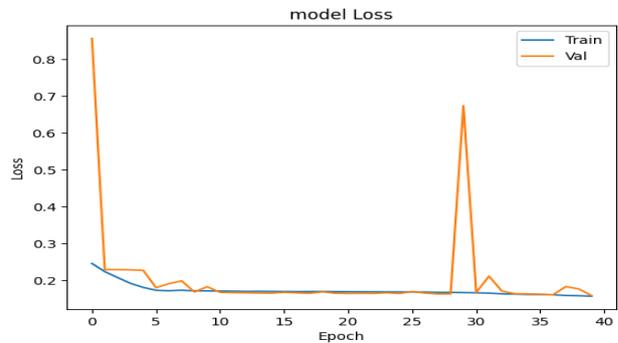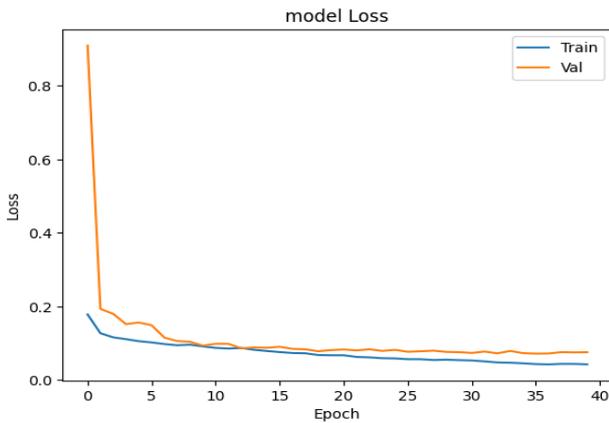

(c) Depth = 2, Layers = 10.
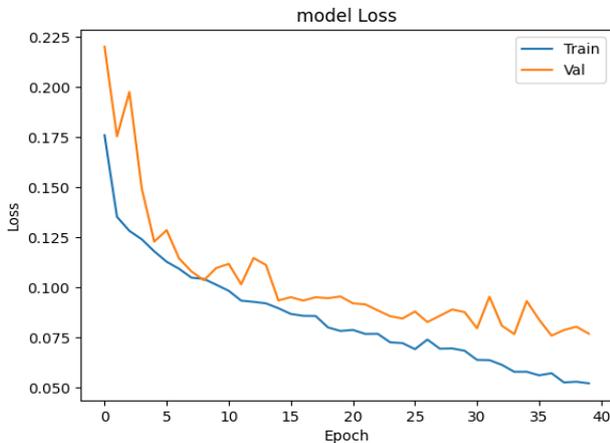

(d) Depth = 3, Layers = 14.
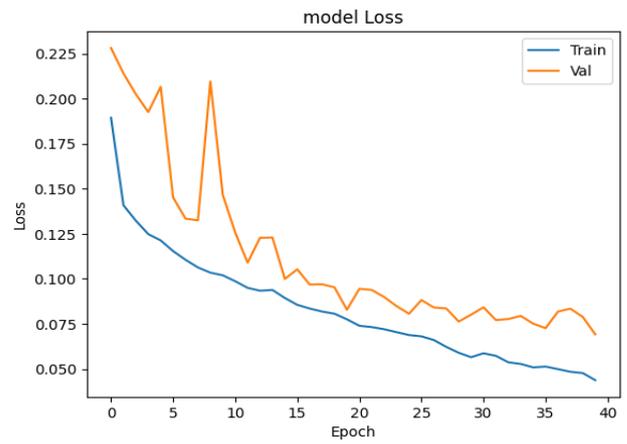
(e) Depth = 4, Layers = 18.

Fig. 3. After model training, Loss Value versus Epoch graphs for [0, 1, 2, 3, 4] architectures are shown for the FCN model. Blue: A training loss. Saffron: Loss of Authenticity/Validation.
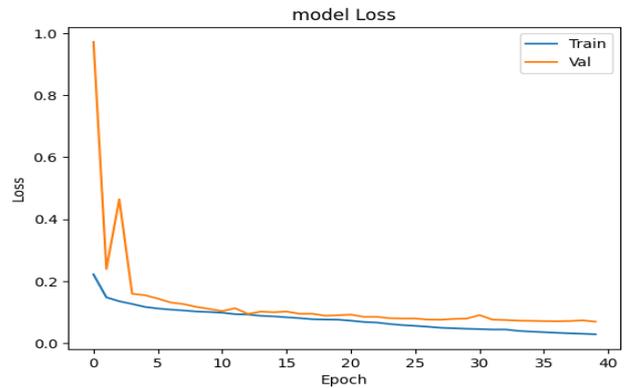

(a) Initial starting filter size K=8, Layers =18, Depth=4.

(b) Initial starting filter size K=16, Layers =18, Depth=4.

(c) Initial starting filter size K=32, Layers =18, Depth=4.

(d) Initial starting filter size K=64, Layers =18, Depth=4.

Fig. 4. Graphs of Loss Value versus Epoch for the FCN model Architecture in Depth 4 with various beginning filter sizes. Blue: Loss in training. Saffron: Loss of Validation.

However, we discover that supplemental data generated by horizontal flipping over, randomized shear, width, height, and zooming shifts performs well for pillar segmentation tasks. According to archaeologists' expert view, pillars are thought to have a diversity of shapes, structures, and orientations but a similar general appearance to neighboring tissues, therefore this discovery makes sense. As a result, we can show that an FCN model cannot be generalized without a specific to a domain data augmentation procedure. The data modifications for augmentation are performed quickly during training, which reduces storage issues. The suggested model can only be trained for 200 epochs (40 epochs/depth), after which no additional changes to the loss function are seen.

Table 3. Designs for architectures using different filters size in depth 4.

Sl.No	Initial Filter Size ($\omega \times \eta \times k$)	No of Parameters	Time elapsed for 40 epochs (hh:mm:ss)
1.	08	491,873	00:42:23
2.	16	1,964,097	00:43:58
3.	32	7,849,601	00:56:14
4.	64	31,384,833	01:13:13

Fig. 3 depicts the training process for several FCN model designs. Depth 0, Depth 1, Depth 2 & Depth 3 models [Ref Fig 3(a)-(d)] have low capacity for learning, and training loss does not considerably diminish. The receptive fields of the Depth 4 model [see Fig. 3(e)] can be further studied. The receptive field of a convolution filter is the portion of the source input image that it effectively covers [34]. If the convolution filter size is maintained across the network, a filter's receptive field expands after each pooling operation. This is due to the fact that although the image's resolution is decreased, the size of the filter does not change. Given that depth 4 models have a larger receptive field and can take more features into account, we select the one with the lowest training loss[see loss column in Table 4]. The loss function's turnover rate change in the Depth 4 model, however, exhibits certain over-fitting trends. By altering a measure of how many filters are in each layer, while maintaining this Depth 4, it is possible to change the number of parameters and lessen over-fitting. Fig. 4 depicts the training procedures for the various architectures created at Depth 4 by altering the beginning filter size, and Table 3 lists the number of parameters in each design.

We see that for both k = 32 and k = 64 initial filter sizes, low training losses are attained with significantly fewer over-fitting tendencies. As a result, the model with a greater number of parameters is chosen as the final design because of its stronger generalizability and fewer over-fitting trends. The proposed FCN model with Depth=4 can be trained in an average of 110 seconds per epoch (with a total of 1055 samples in the training set).

4. Results and Analysis

The suggested approach's segmentation results are compared to manually compiled ground truth (GT), which is supplied with dataset images and verified by a team of archaeology specialists. To determine False Negatives , False Positives, and True Positives, we performed pixel-wise analysis. Pixels that the algorithm incorrectly identified as pillars are known as false positives (FP), True Positives (TP), on the other hand, are described as true pillar pixels found using the approach and False Negatives (FN) are true pillar pixels that the algorithm neglects to detect. For each of the pillar styles in the test dataset, precision and recall measures are calculated against GT. The metrics for recall and precision are as follows:

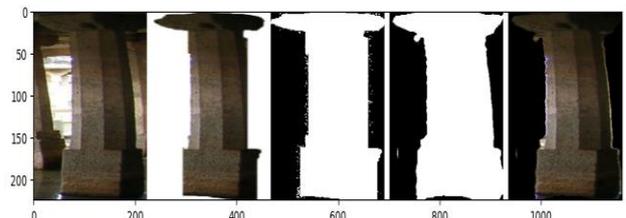$$\text{Precision} = \frac{TP}{TP + FP},$$
$$\text{Recall} = \frac{TP}{TP + FN} \qquad (4)$$

To assess the correlation accuracy of the segmentation, the Dice coefficient is calculated between the algorithm's segmented findings (Detected) and the ground truths (GTs) for each test set. The formula for the dice coefficient is [35]:
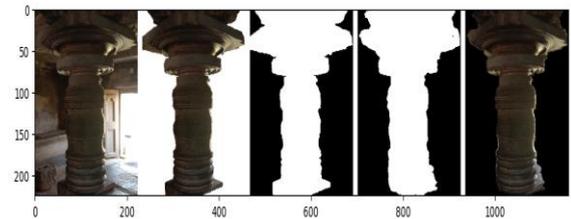
$$\text{Dice coefficient} = 2 \text{ x } \frac{|\text{Detected} \cap \text{GT}|}{|\text{Detected}| + |\text{GT}|} \qquad (5)$$
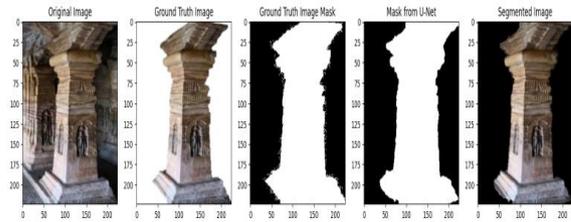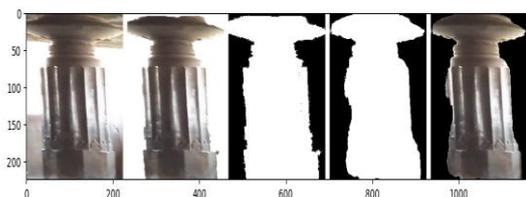

(a) Bhramakantha

(b) Vishnukantha

(c) Rudrakantha

(d) Bhanukantha

(e) Soumyakantha

Fig. 5. Results of various pillar type segmentation of the designed FCN network.

Fig 5 shows the results of proposed FCN network segmentation on different pillar types. The first pillar of these figures displays original pillar images of various types, the second pillar displays a GTs Image, the third pillar displays a GTs Image mask, the fourth pillar displays a mask from the proposed FCN U-Net model, and the last pillar displays the segmentation result of the proposed FCN U-Net model, respectively. The results are shown for each of the five various pillar styles in Figs. 5(a) Bhramakantha, 5(b) Vishnukantha, 5(c) Rudrakantha, 5(d) Bhanukantha, and 5(e) Soumyakantha, respectively.

Table 4. Mean (standard deviation) of recall and precision using the presented method at distinct depths.

Sl.No	Depth	Evaluation parameters			
		Loss	Dice	Recall	Precision
1.	Depth 0	0.2147	0.7849	0.9598	0.6606
2.	Depth 1	0.1985	0.8010	0.9519	0.6878
3.	Depth 2	0.1630	0.8369	0.9525	0.7473
4.	Depth 3	0.1048	0.8946	0.9271	0.8657
5.	Depth 4	**0.0715**	**0.9284**	**0.9698**	**0.9200**

Table 4 displays the suggested method's Dice coefficient , mean precision, recall, and loss results. It is evident that the suggested strategy provided the highest Dice coefficient-0.9284, recall-0.9698, precision-0.9200 and best value for loss-0.0715 obtained from depth4.

Table 5. Comparison of the suggested method's mean (standard deviation) DICE coefficient with the outcomes of the pillar segmentation challenge on all types/styles of pillars (test set).

Sl.No	Methods Name	DICE Coefficient	Method Type
1.	Threshold [17]	0.72	Statistical Method
2.	K-Means [17]	0.49	Statistical Method
3.	Graph-Cut [17]	0.80	Statistical Method
4.	Proposed Modified U-Net FCN Segmentation Model	0.9284	Convolutional Neural Network

Table 5 displays the Dice coefficient utilising the suggested approach in comparison to the publicly available Pillar Segmentation Challenge results. Based on VOE and DSC metrics, which are used to evaluate the effectiveness of segmentation approaches, regional segmentation is more accurate than the threshold methodology, according to a comparative study on temple structural elements [17].

In this work, a pre-processing module is used to equalize the pillar images across different pillar styles. To further characterise forms and ornamentation severity levels (architecture) structures, The FCN model autonomously gathers data at both the micro and macro levels. As a result, the suggested method performs better than the challenge results that have been reported, with the highest Dice rate of 0.9284 and an improvement of 12% over the outcomes given in Narendra K. S. et al. [17].

5. Conclusion and Future Work

An FCN model-based technique for segmenting pillar areas that is independent of style is presented in this paper. The FCN model is trained specifically for pillar region segmentation using denoised pillar pictures. According to the sensitivity assessments of the model hyper-parameters, deeper networks are better at learning novel features than shallower networks, but higher receptive fields result in more training losses, which could lead to model over-fitting. The recommended ideal Modified U-Net FCN model for pillar segmentation generates a style-independent, reduced receptive field model with initial filter size 8 and Depth 4 that is computationally efficient. Additionally, it is discovered that domain-specific data augmentation techniques enhance model training and

convergence rates. On a dataset of old temple pillars, the suggested model is assessed quantitatively and qualitatively. The results demonstrate that the proposed method effectively segments the pillar by providing a mean Dice rate of 0.92 with the least loss of 0.07 among the five different types/styles.

Future research may focus on segmenting idol tasks or detecting idol boundaries inside pillars by appropriately changing the FCN parameters, objectives, and loss functions. Additionally, the segmented pillar images can be used as input for DCNN models that classify and identify the type/style of pillar.

Acknowledgements

Author contributions

Gurudeva Shastri Hiremath: Conceptualization, Methodology, Software, Field study**,** Data curation, Writing-Original draft preparation, Validation. **Shrinivasa Naik C.L. :** Visualization, Investigation, Writing-Reviewing and Editing. **Narendra Kumar S** : Reviewing and Editing and proofreading.

Conflicts of interest

The authors declare no conflicts of interest.

References

[1] Kenyon, Sir Fredrick. The Bible and Archaeology. Harper & Brothers Publishers, New York, 1940.

[2] Andrew S. Kulikovsky B "The Relevance of Archaeology to the Study of Scripture", App.Sc(Hons), February 10, 1994.

[3] History of Indian Archaeology through Western European scholars [Online].Available:https://en.wikipedia.org/wiki/History_of_Indian_archaeology.

[4] Shweta Vardia, "Building science of Indian temple architecture", Master's thesis, university: Universidade do Minho, 21 July 2008.

[5] Ar.Meenal Kumar, 2017. "Pillars [Stambha]- the supportive elements of hindu temples", International Journal of Current Research 9, (05), 50101-50107.

[6] R. R. N. Senthilkumaran, "A Study on Rough Set Theory for Medical Image Segmentation," International Journal of Recent Trends in Engineering, vol. 2, november 2009.

[7] S. k. A. P. Prasad Dakhole, "Fabric Fault Detection Using Image Processing Matlab," International Journal For Emerging Trends in Engineering and Management Research (IJETEMR), vol. 2, no. 1, 21 january 2016.

[8] C. R. Tippana, "Homogeneous Regions for Image Segmentation Based on Fuzzy," international journal & magazine of engineering, technology, management and research.

[9] Rossella Cossu and Laura Chiappini, "A color image segmentation method as used in the study of ancient monument decay", Elsevier Journal of Culture Heritage 5,2004, pp.385-391.

[10] M. Cheriet, J. N. Said, and C. Y. Suen, "A Recursive Thresholding Technique for Image Segmentation", IEEE Trans. On Image Processing, June 1998, Vol.7, issue. 6, pp. 918-921.

[11] F. C. Monteiro and A. Campilho, "Watershed framework to region-based image segmentation," in Proc. International Conference on Pattern Recognition, ICPR 19th, pp. 1-4, 2008.

[12] M. Hameed, M. Sharif, M. Raza, S. W. Haider, and M. Iqbal,"Framework for the comparison of classifiers for medical image segmentation with transform and moment based features," Research Journal of Recent Sciences, vol. 2277, p. 2502, 2012.

[13] S. Zhu, X. Xia, Q. Zhang, and K. Belloulata, "An image segmentation algorithm in image processing based on threshold segmentation," in Proc. Third International IEEE Conference on Signal-Image Technologies and Internet-Based System, SITIS'0., pp. 673-678, 2007.

[14] Arti Taneja, Dr. Priya Ranjan , Dr.Amit Ujjlayan, "A Performance Study of Image Segmentation Techniques" IEEE, 2015.

[15] Maria Mercede Cerimele, Rossella Cossu, "Decay regions segmentation from color images of ancient monuments using fast marching method" Elsevier Journal of Cultural Heritage, 8 (2007) ,pp 170-175.

[16] A. Masiero, A. Guarnieri, F. Pirotti, A. Vettore, "Semi-Automated Detection Of Surface Degradation On Bridges Based On A Level Set Method " The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences,

Volume XL-3/W3, 2015 ,ISPRS Geospatial Week 2015, 28 Sep – 03 Oct 2015, La Grande Motte, France.

[17] Narendra K S, Srinivas N & Gurudeva S H, " Comparative Study on Temple Structural Element Segmentation using Different Segmentation Techniques", I. J. Engineering and Manufacturing, 2023, 2, 32-39.

[18] K. Kamnitsas, L. Chen, C. Ledig, D. Rueckert, and B. Glocker, "Multi- scale 3D convolutional neural networks for lesion segmentation in brain MRI," Ischemic Stroke Lesion Segmentation, vol. 13, pp. 13–16, 2015.

[19] Y. Lequan, H. Chen, Q. Dou, J. Qin, and P. A. Heng, "Automated melanoma recognition in dermoscopy images via very deep residual net- works," IEEE Trans. Med. Imag., vol. 36, no. 4, pp. 994–1004, Apr. 2016.

[20] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervention., 2015, pp. 234–241.

[21] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in Proc. IEEE Conf. Comput. Vision Pattern Recognit., 2015, pp. 3431–3440.

[22] F. Milletari, N. Navab, and S.-A. Ahmadi, "V-net: Fully convolutional neural networks for volumetric medical image segmentation," in Proc. 2016 4th Int. Conf. 3-D Vision, 2016, pp. 565–571.

[23] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in Proc. IEEE Conf. Comput. Vision Pattern Recognit., 2016, pp. 770–778.

[24] Jonathan Masci, Jesus Angulo, and Jurgen Schmidhuber , "A Learning Framework for Morphological Operators using Counter-Harmonic Mean" Conference Paper · December 2012, DOI: 10.1007/978-3-642-38294-9_28.

[25] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in Proc. IEEE Conf. Comput. Vision Pattern Recognit., 2015, pp. 3431–3440.

[26] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervention., 2015, pp. 234–241.

[27] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," arXiv:1409.1556, to be published.

[28] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in Proc. Int. Conf. Mach. Learn., 2015, pp. 448–456.

[29] Ancient Temple Pillar Dataset Prepared by us and used in this research.[Online].Available:https://www.kaggle.com/datasets/devguruap4u/ancient-temple-pillar-images-dataset.

[30] Ancient Temple Pillar Dataset Prepared by us and used in this research.[Online].Available:https://jnnce.ac.in/Temple DataSets/GURUDEV%20%20Description_of_KU-UBDTCE_JNNCE_Temple_Pillar_Database.pdf.

[31] F. Chollet et al., "Keras," GitHub, 20156. [Online]. Available: https:// github.com/keras-team/keras

[32] R. Kohavi et al., "A study of cross-validation and bootstrap for accuracy estimation and model selection," in Proc. 14th Int. Joint Conf. Artif. Intell., Stanford, CA, USA, 1995, vol. 14, pp. 1137–1145.

[33] K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: Sur- passing human-level performance on imagenet classification," in Proc. IEEE Int. Conf. Comput. Vision, 2015, pp. 1026–1034.

[34] W. Luo, Y. Li, R. Urtasun, and R. Zemel, "Understanding the effective receptive field in deep convolutional neural networks," in Proc. 30th Int. Conf. Neural Inf. Process. Syst., 2016, pp. 4898–4906.

[35] L. R. Dice, "Measures of the amount of ecologic association between species," Ecology, vol. 26, no. 3, pp. 297–302, 1945.

[36] Prof. Nitin Sherje. (2017). Phase Shifters with Tunable Reflective Method Using Inductive Coupled Lines. International Journal of New Practices in Management and Engineering, 6(01), 08 - 13. Retrieved from http://ijnpme.org/index.php/IJNPME/article/view/50

[37] Pasha, M. J ., Sreenivasulu, K ., Ramani, B. R ., Sunitha, M. J ., Swetha, K., & Samunnisa, K . (2023). Solid Waste Supervision System based on Heuristic Algorithmic approach and Internet of Things . International Journal on Recent and Innovation Trends in Computing and Communication, 11(1s), 64–70. https://doi.org/10.17762/ijritcc.v11i1s.6000

[38] Dhanikonda, S. R., Sowjanya, P., Ramanaiah, M. L., Joshi, R., Krishna Mohan, B. H., Dhabliya, D., & Raja, N. K. (2022). An efficient deep learning model with interrelated tagging prototype with segmentation for telugu optical character recognition. Scientific Programming, 2022 doi:10.1155/2022/1059004