International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

An Optimized Model for the Segmentation of the Ancient Temple Vimanas using FCN Network

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

Submitted: 10/01/2024 **Revised**: 15/02/2024 **Accepted**: 24/02/2024

Abstract*:* An extensive collection of artifacts, antiquities that are historically and archaeologically significant monuments is housed in the Indian state of Karnataka. Tradition and culture are intricately linked. Karnataka boasts a multitude of Neolithic and Megalithic structures, which have withstood the test of time for millennia. These architectural marvels are remnants of esteemed ruling dynasties. They possess unique wonders characterized by their distinctive style, inherent sculptural and architectural qualities, technical prowess, vastness, and grandeur. Nevertheless, the current generation is ill-prepared to extract archaeological knowledge pertaining to empires or reigning dynasties of these ancient Karnataka temples under the instruction of archaeologists. Therefore, it is necessary to adopt a novel method to effectively deliver this vital information to the contemporary age through a suitable platform. Archaeologists have numerous intricate challenges due to the absence of reliable digital techniques for automatically segmenting Vimana. Automated segmentation of Vimana poses challenges due to the variability in image acquisition, intricate architectural designs, noise, time difficulties, and photographic artifacts. As per our knowledge techniques for segmentation have not been proposed in the literature for vimana segmentation. Our work introduces a optimized fully convolutional network (FCN) model designed specifically for the automated segmentation of Vimana. The suggested approach mitigates the variability of image noise and trains Fully Convolutional Network (FCN) models using images from our custom dataset. Additionally, it has been demonstrated that employing appropriate data augmentation and model hyper-parameterization effectively mitigates over-fitting in the context of vimana area segmentation. The proposed methodology is evaluated using the test dataset, attaining a rate of recall of 0.9302 and a precision rate of 0.8977. The recommended method outperforms four other methods with lower depths in the segmentation problem, earning a Dice correlation coefficient of 0.8894 & with very min loss of around 0.1106. Finally a comparison of same methods with & without edge-smoothing is carried out. An improvement of 12%, 28% is achieved in DICE & PRECISION by an optimized FCN(U-Net) for the segmentation of vimana.

Keywords: *Archaeology; segmentation; vimana; Fully Convolutional Network (FCN); hyper parameters; recall; precision; Dice correlation coefficient.*

1. Introduction

The captivating state of Karnataka in India boasts a plethora of historically and archaeologically significant ancient sites that are truly awe-inspiring. The intertwining of culture and tradition is evident in its essence. Throughout the annals of time, a myriad of awe-inspiring historical monuments have stood the test of centuries, meticulously crafted by the visionary empires and illustrious governing dynasties of yore. The hallmark of their architectural approach lies in their unique methodology of spatial organization, characterized by their distinct style, intrinsic sculptural elements, innovative use of architectural technology, grandiose proportions, and sheer magnitude.

The illustrious and prominent ruling lineages of Karnataka encompassed the esteemed Gangas, Kadambas, Chalukyas of Badami and Hoysalas, Rashstrakutas, Kalayana, as well as the distinguished sultanate dynasties including the Barid Shahis, Adil Shahis, Bahmanies, the esteemed rulers of Vijayanagara, and the revered Wodeyars of Mysore, among an array of other notable entities. During the construction phase, it is noteworthy that these temples, such as Dravida, Nagara, and Vesara, were meticulously crafted in accordance with their respective architectural styles. In the contemporary era, there exists a compelling need for the broader populace as well as scholarly practitioners in the field of archaeology to engage in the extraction of archaeological wisdom by means of digitalization. An exemplary instance entails the exploration of a historic Karnataka temple. Hence, it became imperative to devise a novel approach that would effectively disseminate this crucial information within a contemporary societal framework. This research endeavor shall serve as a digitized instrument for future archaeologists (research scholars) within the realm of archaeology, specifically for the meticulous undertaking of temple field surveys. It shall operate with utmost efficiency, taking into account a multitude of factors such as cost, time, and precision. This endeavor shall serve as a guiding beacon, illuminating the path for esteemed archaeologists in their noble pursuit of temple restoration. By unraveling the intricate tapestry of architectural design choices made in times past, we

1 J.N.N College of Engineering, Shivamoga-577204, Visvesvaraya Technological University, Belagavi-590018, INDIA.
ORCID ID : 0000-0001-6466-279X
2U.B.D.T College of Engineering, Davanagere-577004, Visvesvaraya Technological University, Belagavi-590018, INDIA.
ORCID ID : 0000-0001-9019-1733
3St. Joseph Engineering College, Mangaluru-575028 , Visvesvaraya Technological University, Belagavi-590018, INDIA.
ORCID ID : 0000-0001-5968-2841
**Corresponding Author*
Email: devguruap4u@gmail.com/gurudevah@sjec.ac.in

shall empower these scholars to navigate the restoration processes with unparalleled insight and wisdom.

Currently, archaeologists are addressing the aforementioned matters through manual means, as there has been no implementation of digitalization in relation to the temples of India. In this context, it is imperative to incorporate an exceedingly efficient image segmentation technique for the purpose of conducting image segmentation as an integral component of the image pre-processing phase. This particular step primarily concentrates on the identification and delineation of the temple's architectural elements, with a specific emphasis on the vimana structure.

Segmentation of the image stands as an imperative technique within the realm of image processing. The pre-processing stage holds significant importance within the realms of image analysis , computer vision and pattern recognition [1]. Image segmentation is a sophisticated methodology employed to meticulously dissect a digital image into distinct components, each comprising a unique collection of pixels. In the realm of architectural design, the utilization of image segmentation is a common practice to discern and delineate various objects and boundaries within images. This technique involves the meticulous identification of lines, curves, and other visual elements, allowing for a comprehensive understanding and analysis of the image's composition. When an image undergoes the process of segmentation, it has the potential to yield a collection of segments that encompass the entirety of the image or a series of extracted image outlines achieved through edge detection. In terms of various attributes or calculated characteristics, such as hue, saturation, or pattern, there exists a cohesive connection among all the pixels within a given region. The user's text is too short to rewrite in the style of an architectural designer. Please provide more information In relation to the identical attributes, the neighboring regions exhibit notable disparities. Image segmentation is a pivotal technique employed in various industries, including medical imaging, face recognition, digital libraries, computer vision, image processing, and picture and retrieval of video [2].

Image segmentation methods encompass a wide range of techniques, including clustering-based, edge-based, feature-based, thresholding, and artificial neural networks based segmentation. Segmenting images can be effectively achieved by grouping of all the photographic images together [3]. The segmentation process's efficiency was enhanced through the utilization of various methods, specifically K-means clustering, level set, active contour and Fuzzy clustering, as described by ArtiTaneja et al. [4]. The author thoroughly examines the performance of algorithms for image segmentation. This level set methodology has two distinct levels: texture-based and intensity-based segmentation of images. The integration of both intensity and texture-based image segmentation yields superior outcomes compared to conventional techniques.

The authors in Maria Mercede et al [5] present a novel color image segmentation strategy using the fast marching numerical method. This technique is exclusively used to the boundaries of the specific area being analyzed. Furthermore, they introduced a comprehensive concept, namely, the potential extraction of decay zones from the entire image; these regions were spatially separate but had identical colorimeter values.In this study, A. Masiero et al [6] provide a detailed analysis of the identification of deterioration on building surfaces. They propose the use of an advanced technique for image segmentation utilizing a level sets technique. This method of segmentation reduces the cost of monitoring and guarantees that the eventual output is unbiased.

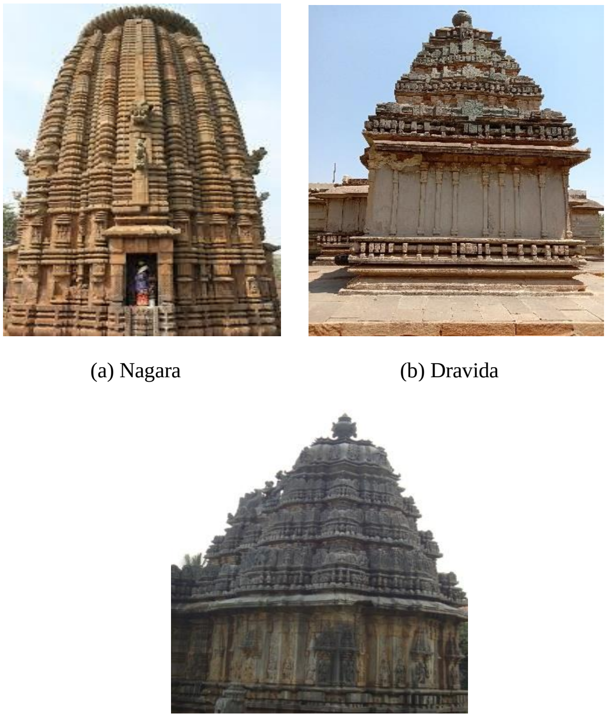
K. Kamnitsas et al. [7,8] provide a detailed account of the encouraging outcomes achieved through the utilization of convolutional neural networks (CNNs) for segmenting biological pictures, in comparison to older methods. CNNs, or Convolutional Neural Networks, are a distinct neural network architecture that employ iterative Conducting convolution techniques to obtain specific features relevant regarding the segmentation task, it involves processing an input image. The neurons' learnable biases and weights provide convolution filters. Different configurations of these filters can be utilized to create architectural solutions for specific segmentation challenges.

The U-Net architecture, created by Ronneberger et al, is a widely used CNN model for various clinical segmentation of images applications [9]. The Fully convolutional network (FCN), U-Net gained popularity through its introduction by Long et al [10]. If an input image is provided, the output will be the corresponding segmentation mask. Milletari and other authors in [11] modified the model belonging to them by utilizing stacks of 3D images as input, so advancing this kind of approach. The ResNet discusses the idea of residual connections design by X. Zhang et al in [12], was also introduced by them.The choice to utilize FCN in the context of architecture-independent vimana segmentation challenges is motivated by the inherent ability of CNNs to engage in self-learning processes and abstract learning, enabling them to discern subtle spatial distinctions. The segmentation of vimanas within ancient temples has not yet been accomplished through the utilization of Fully Convolutional Networks (FCNs).

In order to effectively classify the diverse array of vimanas & discern the different styles of architecture adopted in their meticulous construction, archaeologists have the capacity to identify the specific structural characteristics of these vimanas through the utilization of automated segmentation methods. This will aid them in making more informed decisions during the process of temple reconstruction. In this endeavor, we have opted to employ CNN-based segmentation methodologies as our approach. Our meticulous examination of the existing literature has demonstrated that these methods have exhibited superior performance compared to other segmentation techniques.

In the contemporary architectural discourse, a plethora of cutting-edge convolutional neural network (CNN) segmentation methodologies have been proffered. In accordance with our current understanding, the selection of acquisition and visualization-specific parameters is executed through established methodologies that rely on human intervention. The dynamic nature of the noise level and the diverse range of pixel intensities observed in the vimana images pose a considerable challenge when it comes to the segmentation of vimanas. Figure 1 showcases a collection of visually captivating vimanas, captured with meticulous attention to the intricate architectural forms they embody. The proposed approach effectively tackles the challenges that arise during the development of an automated, self-sufficient vimana segmentation method.

This work presents two pivotal contributions. Presented here is a highly versatile and innovative Fully Convolutional Network (FCN) model, meticulously designed to effortlessly segment vimanas from images, even when they are accompanied by extraneous backgrounds that are not desired. In relation to the objectives of vimana area segmentation, we thoroughly examine how sensitive some model parameters, like the number of layers, filter size, and network depth, which are the fundamental dimensions. Additionally, we assess the training accuracy, testing accuracy, and validation accuracy as crucial metrics. In our analysis, we have discovered that by skillfully parameterizing the model, we can reach a commendable recall-rate of 0.9302, while simultaneously upholding a precision-rate of 0.8977 across a different range of vimana types. This performance surpasses that of existing approaches, which boast a superior Dice coefficient of 0.8894.



(a) Nagara

(b) Dravida

(c) Vesara

Fig 1. Different Styles of Vimanas (a), (b) and (c) Based on ancient architecture.

The remaining components of this research work are re-organized in the following section explained manner. Section II delves into the intricate realm of data-sets and presents the proposed FCN methodology for the purpose of auto-segmentation. The elucidation of the experimental setups can be found in Section III. The presentation and discussion of the conducted experimental results can be explained in Section IV. Section V, in its culmination, serves as the platform for drawing insightful conclusions and engaging in meaningful discussions.

Archaeologists can utilize automated segmentation methods to determine the structural properties of temple vimanas and classify them according to their architectural styles. This will assist them in making more informed judgments during the process of reconstructing the temple. In this study, we opted to utilize CNN-based segmentation methods due to the findings of our literature review, which demonstrated their superior performance compared to other segmentation techniques. Several CNN segmentation approaches have been proposed in recent literature. As far as we know, the selection of acquisition and visualization-specific parameters is currently done using approaches that rely on human intervention. The vimana segmentation process is challenging due to the fluctuating noise level and varying pixel brightness observed in the vimana photos recorded at different times. Figure 1 exhibits images of Vimanas captured according to the degree of intricacy in their shapes and embellishments. The proposed methodology tackles the challenges that arise in developing an automated independent vimana segmentation algorithm.

This paper makes two significant contributions. The initial demonstration showcases a versatile and state-of-the-art FCN model capable of automatically distinguishing vimanas from pictures that contain both vimanas and unwanted backgrounds. We examine the impact of model parameters, such as core dimensions and layer count, on the objectives of vimana area segmentation. Upon examination, we find that a well-configured model may attain a recall rate of 0.9698, surpassing state-of-the-art techniques, while simultaneously upholding a precision rate of 0.9284 across various types of vimanas. An optimized FCN (U-Net) achieves a significant improvement of 12% and 28% in DICE and PRECISION, respectively, for the segmentation of vimana.

The remaining sections of the work are structured as follows. The methods and data are discussed in Section II. The experimental circumstances are elucidated in Section III. Section IV shows and analyzes the results obtained from the experiment. The debates are concluded in Section V.

2. Data-sets and proposed FCN methodology

Experiments have been conducted on our internally curated data-set of ancient-temple vimanas to evaluate the efficacy of the new proposed method. Within the realm of architectural severity, this curated assortment of ancient temple vimanas showcases three distinct typologies. The dataset has been graciously made available exclusively only for research purposes through our esteemed J.N.N.C.E college web-site and the esteemed Kaggle web-portal [13, 14].The diverse typologies of vimanas have been duly validated by esteemed archaeologists with their expertise. The dataset comprises a collection of meticulously curated images, each of which is accompanied by an expertly assigned label denoting distinct architectural styles, categorized according to the severity levels of architectural elements. These styles include the illustrious Nagara, the elegant Dravida, and the refined Vesara, as visually depicted in Figure 1. This ancient-temple vimana dataset has been partitioned into distinct training and test subsets. The dataset description is elegantly presented in the exquisite Table 1. The provided visual representation in Figure 1 showcases a meticulously curated assortment of vimana images, each belonging to its respective category.

Pre-processed Vimana images are segmented to exclude the backdrop so that vimana images are localized. An optimized U-Net model which draws inspiration from Ronneberger's U-Net design [9], is employed for the semantic segmentation of vimana.

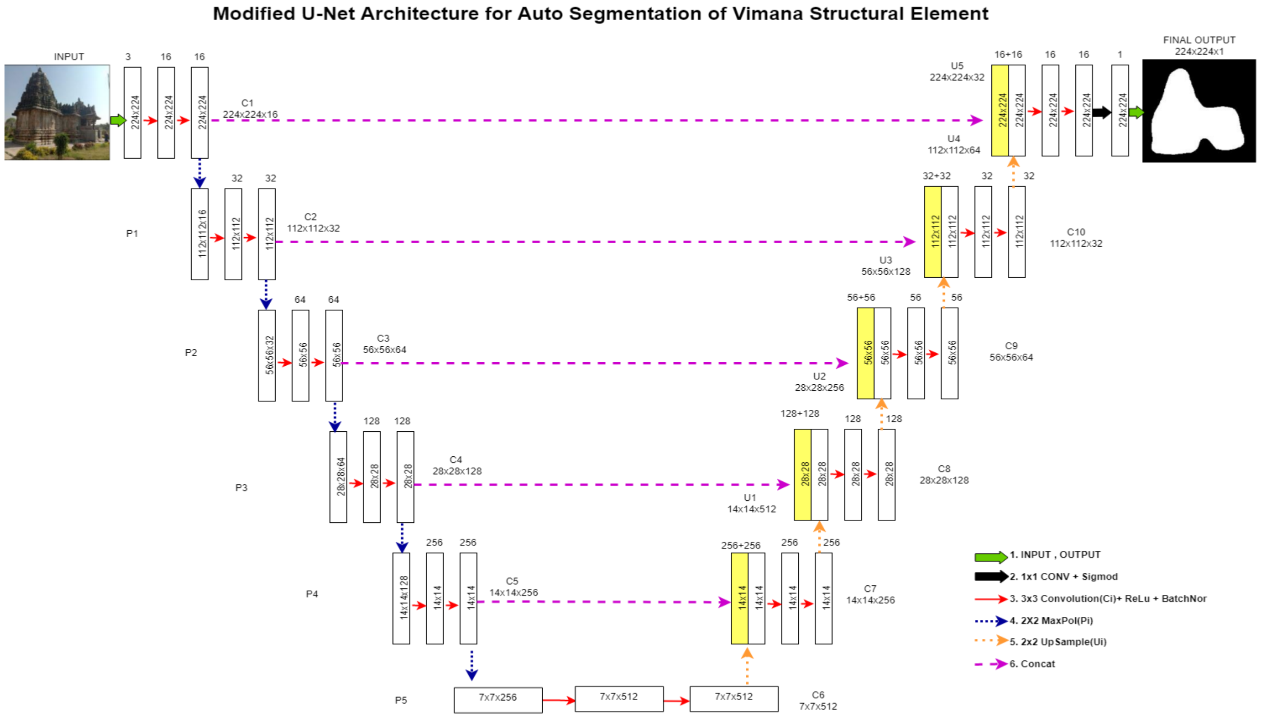


Fig 2. The proposed FCN model consists of 18 convolutional layers, each composed of a network architecture comprising between 32 and 512 kernels.

The suggested model's architecture is depicted in Figure 2. Optimized U-Net model consists of a contracting path (encoder) on the left side, bottleneck at the bottom and a expansive path (decoder) on the right side which makes it an ’U’ shaped architecture. As shown in Figure 2, all the open boxes corresponds to a multi-channel feature map, shaded box in encoder indicates the input layer and shaded box in decoder part represents the copied feature map from the encoder. The number of channels of the feature map were denoted on top of the each box. Proposed U-Net model contains total 23 convolution layers. Encoder: It has five depth levels, with two 3×3 convolutions layers, ReLU activation, Batch-Normalization [16], and a 2 × 2 max pooling operation with stride 2 at the end of each depth level. Number of feature channels in the first depth level is 16 and these feature channels were doubled for each further depth level of encoder as shown in the Figure 2. The operation in the encoder is called down-sampling performed using max pooling layer, in which image size is decreased by half in each depth level of the encoder.

Table 1. Dataset Description

Sl.No	*Vimana Type*	*Train set*	*Valid set*	*Test set*	*No of images per each type of vimanas*
1.	Nagara	607	173	86	253
2.	Dravida	607	173	86	313
3.	Vesara	607	173	86	300
Total Images in the Vimana Ancient Temple Dataset.					866

The bottleneck consists of two 3 × 3 convolution layers, which are then followed by ReLU activation function and Batch-Normalization process. The bottleneck contains a total of 512 feature channels. The Decoder block is designed to reverse the activations of the encoder in order to obtain a probability feature-map that is the equal size as of the original input image. Transposed convolution is employed to provide localization, and this technique is referred to as up-sampling. The decoder unit is structured with multiple depth levels. Each depth level includes a 2 × 2 up-sampling convolution, which augments the dimensions of the feature-map. The up-sampled feature map is then concatenated with the corresponding feature map from the encoder using a skip connection. After the concatenation, there are two 3 × 3 convolution layers, which have the same size feature map as the encoder block depth level. Each convolution layer is followed by a ReLU activation function and a Batch-Normalization operation. The last layer employs a 1 × 1 convolution operation, followed by sigmoid activation, to transform each 16-component feature vector into its matching class label. Table 1 displays the primary distinctions between the proposed optimized U-Net model and Ronneberger's U-Net model [9].

The final result of the [1 x 1] layer is utilized to construct the network's loss function. The binary cross-entropy loss function is used to transform the input onto a probability mapping that has the exact dimensions as the input image. The loss function for an individual input image without any pixels is given in (1).

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

where si is the anticipated binary output for pixel i and ti is the actual binary output (target). The weighted total of the inputs and the final output, y, as well as the sigmoid activation function S, are presented in (2).

$$s_i = \frac{1}{1 + e^{-y_i}}\,, \qquad y_i = \sum_{j=1}^{n} x_j w_{ji} \qquad (2)$$

3. Experimental Investigation

Keras 1.0 integrates the recommended convolutional neural network (CNN) model architecture [31] in Google Colab with GPU support.

3.1. Optimization of FCN Model by Training and Hyper Parameterization.

To achieve the desired results, it is necessary to finely adjust the model hyper-parameters. The key parameters/hyper-parameters that require optimization in the context of FCNs include the weight and bias quantities, layer count, filter/kernel count in every layer, and the learning rate of the model. The ideal amalgamation of these elements characteristics is determined by utilizing the hold-out strategy in grid-search [17]. Ultimately, an evaluation of the trained model is conducted using the test dataset. In order to verify the FCN framework architecture, as outlined in Section 2 (Fig 2), several experiments have been conducted. As stated by reference [9], it is customary to place a layer with maximum pooling or a the deconvolution layer after two convolutional layers. The number of layers in different architectures can vary while yet preserving this characteristic, as stated in reference [9].

Table 2 presents a concise overview of different FCN architectures. Depth corresponds to the quantity of max-pooling layers, parameters correspond to the quantity of weights as well as biases in the network, and layers correspond to the quantity of convolutional layers. Timings for Training, Validation & Testing for different FCN architectures are also shown.

The training sets, validation sets, and test sets consist of 607, 173, and 86 images, respectively, representing each of the three architectural styles. These sets are meticulously curated to assess the experimental performance of the Fully Convolutional Network (FCN) parameters across various design variations. The visual representations are subsequently scaled down utilizing bilinear interpolation to adhere to a standardized resolution of [256X512]. In the realm of vimana segmentation tasks, we have discovered that the incorporation of supplementary data generated through techniques such as horizontal flipping, random shear, height and breadth adjustments, as well as zoom shifts, yields commendable results. This discovery aligns with the notion that vimanas exhibit a diverse array of forms, compositions, and alignments, while still maintaining a cohesive semblance to the surrounding anatomical structures, as per the expert perspective of archaeologists. Upon careful examination, it becomes evident that the incorporation of a domain-specific data augmentation process is of utmost importance in order to bestow upon an FCN model the coveted attribute of generalizability. The architectural design effectively mitigates storage concerns by implementing instantaneous data transformations for augmentation during the training process.

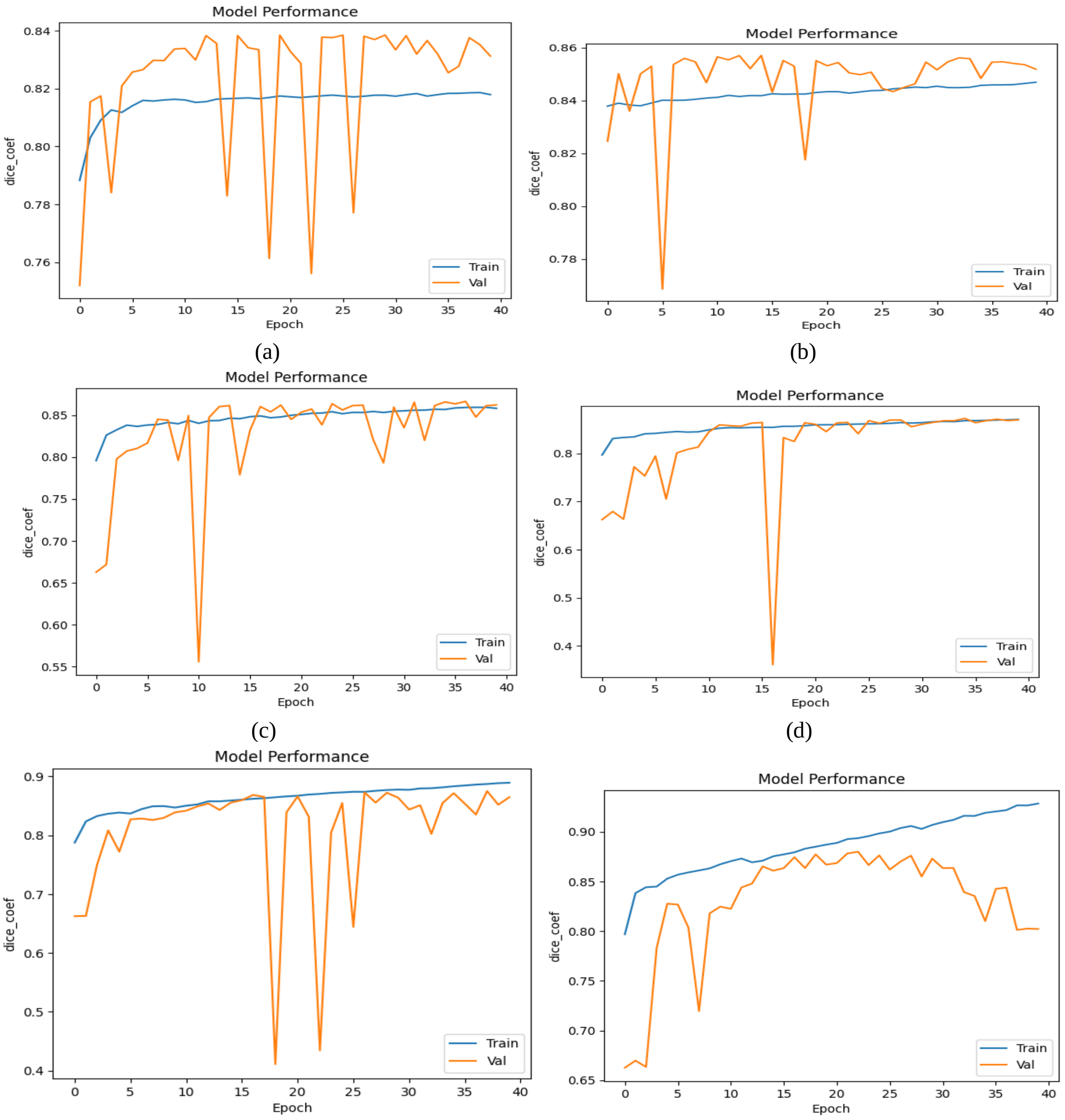
Table 2. Various FCN architectures are generated by altering the number of layers and depths, with a filter size of (ω × η × k) = (3 × 3 × 64).

Sl.No	***No of Layers***	***Depth***	***Total No of Parameters***	***Trainable params***	***Non-trainable params***	***Training Time elapsed (Mins)***	***Validation Time elapsed (ss:ms)***	***Testing Time elapsed (s:ms)***
1	04	0	39169	38913	256	12 Mins	16s 162ms/step	1s 79ms/step
2	06	1	409345	408321	1024	19 Mins	17s 448ms/step	2s 50ms/step
3	10	2	1886977	1884417	2560	16 Mins	25s 659ms/step	1s 115ms/step
4	14	3	7791361	7785729	5632	20 Mins	29s 768ms/step	2s 149ms/step
5	18	4	31396609	31384833	11776	24 Mins	32s 855ms/step	1s 19ms/step
6	22	5	125793025	125768961	24064	34 Mins	44s 1ms/step	2s 358ms/step

The proposed architectural design entails training the model for a total of 240 epochs, with each depth consisting of 40 epochs. Beyond this point, no further modifications to the model's accuracy and loss are observed. Additionally, it is worth noting that a decline in validation accuracy becomes apparent, as depicted in Figure 3(f).

Figure 3 showcases the training and validation procedure for diverse FCN model designs. The architectural design encompasses models of varying depths, ranging from 0 to 5, as depicted in Figure 3(a)-(f). These models exhibit a limited capacity for learning, as evidenced by their inability to significantly reduce training loss and dice-coefficient. Further exploration can be conducted to analyze the receptive fields of the Depth 5 model, as depicted in Figure 3(f). It is worth noting that the achievement of minimal training losses, coupled with significantly reduced over-fitting tendencies, is observed when utilizing a filter size of 64 at the initial stage (k = 64).

Consequently, given its heightened capacity for generalization and diminished susceptibility to over-fitting, the model boasting a greater abundance of parameters, which exhibits the most favorable dice coefficient, is chosen as the ultimate design. The proposed Depth 5 FCN model undergoes an average training duration of 32 seconds and 855 milliseconds per epoch, with a total of 607 samples in the training set.



(a)

(b)

(c)

(d)

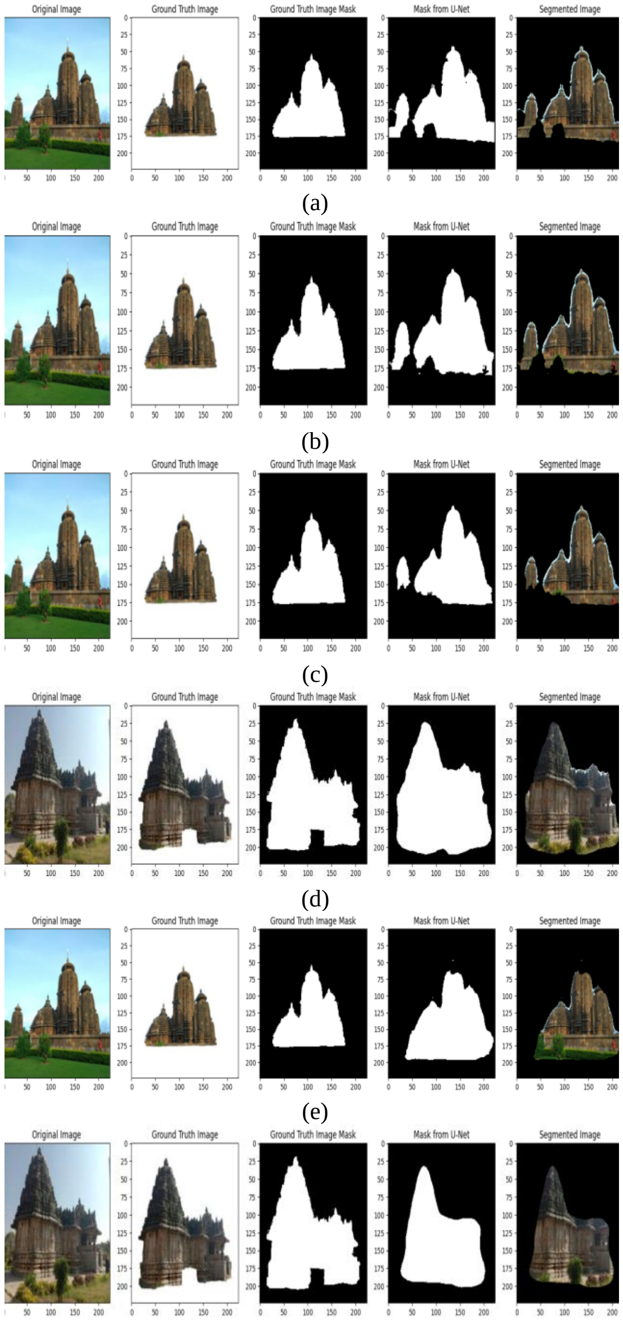
(e)

(f)

Fig 3. Graphs of Model Performance versus Dice-Coefficient Value for the FCN model after model training, depth: [0, 1, 2, 3, 4, 5] architectures. Blue: Accuracy in training. Saffron: Accuracy in Validation. (a)Depth = 0, Layers = 4. (b) Depth = 1, Layers = 6. (c) Depth = 2, Layers = 10. (d) Depth = 3, Layers = 14. (e) Depth = 4, Layers = 18. (f) Depth = 5, Layers = 22.

4. Results Findings and Interpretations

The outcomes of the suggested approach's segmentation are juxtaposed with the meticulously curated ground truth (GT) that accompanies the images of dataset and has been meticulously vetted by a team of esteemed archaeology specialists. In order to ascertain the False Negatives, False Positives, and True Positives a meticulous pixel-wise analysis was conducted. The true positives (TP) correspond to the pixels that are identified as belonging to the object and indeed do belong to the object. On the contrary, the true negatives (TN) refer to the pixels located outside the object, included in the segmentation as well as the actual data (ground truth). The Dice score serves as a performance metric in the realm of image segmentation challenges. In the realm of architectural design, we encounter a distinction between accuracy and its counterpart, where the primary aim is to align the values, as opposed to dice, which not only aligns the value but also the position. To find the dice coefficient, following formula is used:



(a)

(b)

(c)

(d)

(e)

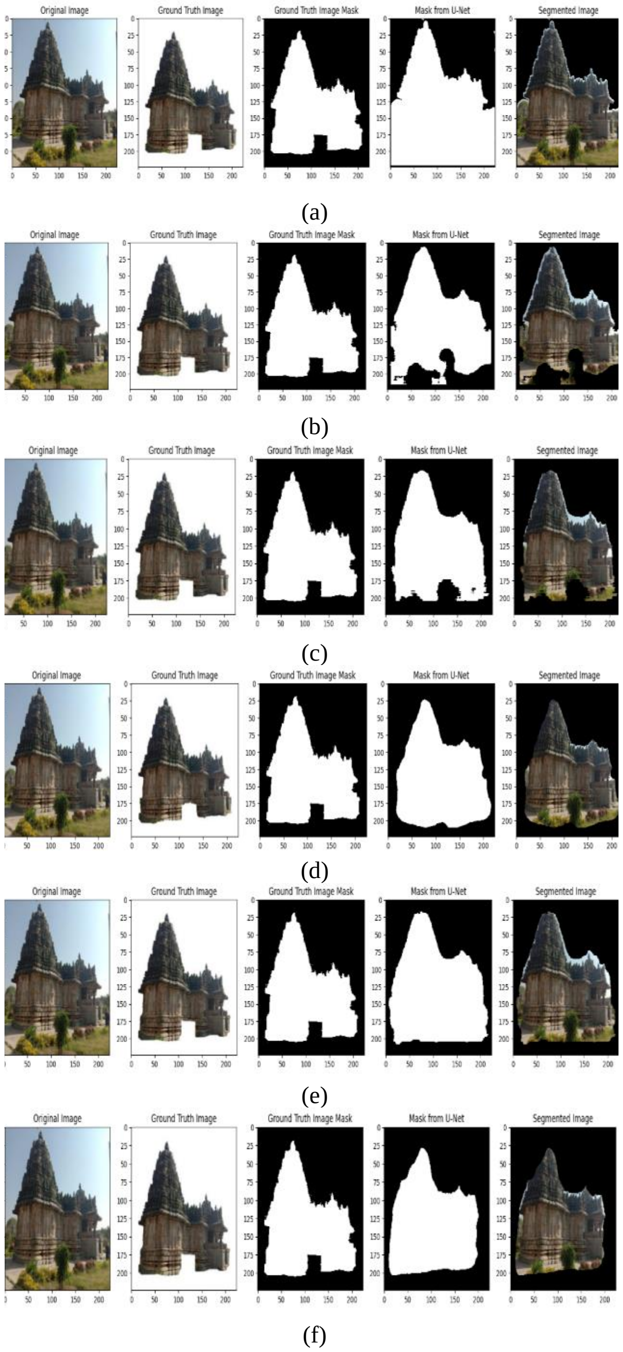
(f)

Fig 4. Results (a) to (f) of proposed optimized different FCN networks segmentation with edge-smoothing on vimana images from depth=0 to 5 w.r.t Figure 3.

$$\text{Dice}_{\text{coefficient}} = 2\text{ x}\frac{|A \cap \text{B}|}{|A| + |\text{B}|} = \frac{2(\text{TP})}{(2\text{TP} + \text{FP} + \text{FN})} \quad (3)$$

Where : A- Detected, B- Ground Truth & FP-False Negative. The following are the precision and recall metrics:

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



(a)

(b)

(c)

(d)

(e)

(f)

Fig 5. Results (a) to (f) of proposed optimized different FCN networks segmentation with-out edge-smoothing on vimana images from depth=0 to 5 w.r.t Figure 3.

Figure 4 displays the outcomes of the segmentation performed by the proposed FCN network on several vimana categories with edge-smoothing. Various varieties of original vimana images are seen in the image on the left of these figures. A GTs Image and its mask are displayed in the second and third columns, respectively. A mask created using the suggested FCN U-Net model is shown in the fourth column, and the segmentation output is shown in the last column. The findings are displayed in figure 4, with (a) to (f) representing the six different FCN structures. These architectures were generated by adjusting the number of layers and depths, while keeping the filter size constant at 64. Figure 3 illustrates this relationship between (a) to (f) and the filter size.

Table 3. Mean (standard deviation) of recall and precision using the presented method at distinct depths with edge-smoothing.

Sl.No	*Depth*	*Evaluation parameters*			
		Loss	*Dice*	*Recall*	*Precision*
1.	Depth 0	0.1679	0.8312	0.9162	0.7549
2.	Depth 1	0.1473	0.8518	0.9189	0.7976
3.	Depth 2	0.1371	0.8621	0.9178	0.8164
4.	Depth 3	0.1296	0.8696	0.8840	0.8598
5.	Depth 4	**0.1106**	**0.8894**	**0.9302**	**0.8977**
6.	Depth 5	0.1976	0.8003	0.7266	0.8652

Table 4. Mean (standard deviation) of recall and precision using the presented method at distinct depths with-out edge-smoothing.

Sl.No	*Depth*	*Evaluation parameters*			
		Loss	*Dice*	*Recall*	*Precision*
1	Depth 0	0.2150	0.7814	0.9162	0.6881
2	Depth 1	0.1872	0.8091	0.8949	0.7452
3	Depth 2	0.1645	0.8325	0.8585	0.8149
4	Depth 3	0.1727	0.8254	0.8619	0.8294
5	Depth 4	**0.1431**	**0.8517**	**0.8725**	**0.8421**
6	Depth 5	0.1502	0.8455	0.8600	0.8404

Figure 5 displays the outcomes of the segmentation performed by the proposed FCN network on several vimana categories with-out edge-smoothing. The findings are displayed in figure 5, with (a) to (f) representing the six different FCN structures with-out edge-smoothing. These architectures were generated by adjusting the number of layers and depths, while keeping the filter size constant at 64.

Table 3 showcases the exquisite manifestation of the suggested methodology with edge-smoothing, wherein the loss, Dice coefficient, recall and mean precision results are elegantly presented. The proposed approach exhibits a remarkable performance, as evidenced by the highest achieved Dice coefficient of 0.8894, recall of 0.9302, precision of 0.8977, and the optimal loss value of 0.1106, all of which were obtained from the depth4 model. Table 4 showcases the exquisite manifestation of the suggested methodology with-out edge-smoothing, wherein the loss, Dice coefficient, recall and mean precision results are 0.1431, 0.8517, 0.8725, 0.8421, all of which were obtained from the depth4 model.

Table 5. Comparison of Original FCN(U-Net) with Optimized FCN(U-Net) for Segmentation.

Sl.No	*Parameters*	*Original*	*Optimized*
1.	Number of Depth Level in encoder and decoder part.	03	04
2.	Filters Size	16	64
3.	Each Depth level in encoder part contains.	Two 3x3 convolutions+ ReLU and a 2x2 max pooling	Two 3x3 convolutions+ ReLU+BatchNormalization and a 2x2 max pooling
4.	Number of feature channels in each depth level.	64,128,256,512	16,32,64
5.	Padding in convolution operation.	Padding is not used. Because of that size of the feature maps will be reduced after each convolution operation.	Padding is used so that size of the feature maps before and after the convolution is same.
6.	Input and Output size	Input: 572*572*1 Output: 388*338*2	Input: 224 * 224*3 Output 224 * 224*1
7.	Data augmentation during the model Used training	Used	Not used
8.	DICE	0.7599	0.8894
9.	RECALL	1.0000	0.9302
10	PRECISION	0.6133	0.8977

Figure 4, Figure 5, Table 3 & Table 4, showcases the results of the suggested methodology with and with-out edge-smoothing. The proposed optimized-FCN exhibits a remarkable performance in segmentation with edge-smoothing rather than the with-out edge-smoothing. This shows how a edge-smoothing plays a very important role in image segmentation.

Table 5 presents a comparison between the original FCN(U-Net) and the optimized FCN(U-Net) for the segmentation of vimana. The table's findings demonstrate the superior performance of the optimized method compared to the existing FCN. An improvement of 12%, 28% is achieved in DICE & PRECISION by an optimized FCN(U-Net) for the segmentation of vimana.

5. Conclusion and future work

This paper presents a technique for segmenting vimana areas using an FCN model-based approach, which is designed to be style-independent. In the realm of architectural design, it has been observed through meticulous investigation of the hyperparameters of the model that deeper networks possess a heightened capacity for acquiring robust features in comparison to shallower networks. This behavior continues until the occurrence of overfitting, wherein the model turns into excessively customized according to training data. It is worth noting that an increase in depth to four levels yields amplified training accuracy and diminished losses. Conversely, a further increase to a depth of five levels leads to elevated losses and reduced training accuracy, potentially resulting in the undesirable consequence of over-fitting the model.

When evaluating the proposed architectural design for the dataset of ancient temple vimanas, a comprehensive assessment is conducted both in terms of quantity and quality. The results indicate that the suggested approach successfully partitions the vimana by achieving a Dice coefficient of 0.8894, recall of 0.9302, precision of 0.8977, and a minimal loss value of 0.1106. These metrics were obtained by evaluating the method on vimana using six different FCN Segmentation network Models with edge-smoothing. The results show how a edge-smoothing plays a very important role in image segmentation. Equally an improvement of 12%, 28% is achieved in DICE & PRECISION by an optimized FCN(U-Net) for the segmentation of vimana.

Future study could prioritize the segmentation of idol tasks or the identification of idol boundaries within vimana by adjusting the FCN parameters, objectives, and loss functions accordingly. Moreover, the segmented vimana pictures can serve as input for DCNN models that categorize and recognize the specific type or style of Vimana architecture.

Acknowledgements

We express our gratitude for the assistance and cooperation extended by **Dr. S. Y. Somashekhar, Dr. Ramesh Nayak, Dr. C. S. Vasudevan, and Dr. Vasudev Badiger**, esteemed researchers and specialists in the fields of Cultural History Archaeology, Art, and Architecture, who are affiliated with the Department of Archaeology & Ancient History at Kannada University, Hampi, India. We express our gratitude to the these individuals who provided valuable assistance in gathering datasets, categorizing them, and sharing their extensive knowledge in the field of archaeology for this research endeavor. We express our equal gratitude to the **Archaeological Survey of India** for granting us permission to capture pictures and for the assistance provided by their employees during the temple field visit.

Author contributions

Mr. Narendra Kumar S contributed to the conceptualization, methodology, software development, field investigation, data curation, original draft preparation, and validation of the project. **Dr. Shrinivasa Naik C.L.** specializes in the tasks of reviewing, editing, and proofreading. **Dr. Gurudeva Shastri Hiremath** specializes in the fields of visualization, investigation, writing-reviewing, and editing.

Conflicts of interest

The authors declare no conflicts of interest.

References

[1] Senthilkumaran, N., & Rajesh, R. (2009). A study on rough set theory for medical image segmentation. *International journal of recent trends in Engineering*, *2*(2), 236.

[2] Panwar, P., Gopal, G., & Kumar, R. (2016). Image Segmentation using K-means clustering and Thresholding. *Image*, *3*(05), 1787-1793.

[3] Panda, S. (2015). Color image segmentation using K-means clustering and thresholding technique. *Interntional journal of ESC*, 1132-1136.

[4] Taneja, A., Ranjan, P., & Ujjlayan, A. (2015, September). A performance study of image segmentation techniques. In *2015 4th International Conference on Reliability, Infocom Technologies and Optimization (ICRITO)(Trends and Future Directions)* (pp. 1-6). IEEE.

[5] Cerimele, M. M., & Cossu, R. (2007). Decay regions segmentation from color images of ancient monuments using fast marching method. *Journal of Cultural Heritage*, *8*(2), 170-175.

[6] Masiero, A., Guarnieri, A., Pirotti, F., & Vettore, A. (2015). 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*, *40*, 15-21.

[7] Kamnitsas, K., Chen, L., Ledig, C., Rueckert, D., & Glocker, B. (2015). Multi-scale 3D convolutional neural networks for lesion segmentation in brain MRI. *Ischemic stroke lesion segmentation*, *13*, 46.

[8] Yu, L., Chen, H., Dou, Q., Qin, J., & Heng, P. A. (2016). Automated melanoma recognition in dermoscopy images via very deep residual networks. *IEEE transactions on medical imaging*, *36*(4), 994-1004.

[9] Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18* (pp. 234-241). Springer International Publishing.

[10] 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).

[11] Milletari, F., Navab, N., & Ahmadi, S. A. (2016, October). V-net: Fully convolutional neural networks for volumetric medical image segmentation. In *2016 fourth international conference on 3D vision (3DV)* (pp. 565-571). Ieee.

[12] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 770-778).

[13] https://www.kaggle.com/datasets/narendrakumarsubdtce/ancient-temple-vimana-images-dataset Ancient Temple Vimana images Dataset .

[14] https://jnnce.ac.in/TempleDataSets/NARENDRA%20Description_of_KU-UBDTCE-JNNCE_Temple_Vimana_Dataset.pdf Ancient Temple Vimana images Dataset.

[15] Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.

[16] Ioffe, S., & Szegedy, C. (2015, June). Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *International conference on machine learning* (pp. 448-456). pmlr.

[17] Kohavi, R. (1995, August). A study of cross-validation and bootstrap for accuracy estimation and model selection. In *Ijcai* (Vol. 14, No. 2, pp. 1137-1145).