Research Article

The Use of CNN Network in the Generation of Clear Line Contours of Thangka Art

Xinyun Zhang

St. Mark's School, 25 Marlboro Rd, Southborough, MA, 01772, United States

ABSTRACT

Thangka has been a crucial aspect of Buddhism for thousands of years. It acts as a visualization of the myths and legends that shaped Buddhism and attracted many devotees to worship. Furthermore, Thangka also serves as a reflection of the cultural, political and social aspect of Tibetan society in history. However, the paintings, especially the outlines, became extremely fragile after years of harsh environments. I utilized Deep learning, specifically the Convolutional neural network (CNN) as the main method for processing data. After numerous training the training loss gradually approaches 0, which means that the model no longer needs additional training. Under the CNN network I used max pooling which discards trivial information, rectified linear (ReLU) activation function, residential Network (ResNet), and other methods in this research to achieve a clear outline of Thangka. Through episodes of convolution and ReLU function, the image generated becomes gradually more defined in layer 1. However, as the image undergoes continuous process, it gradually becomes more abstract. Thus, the best results could be found in layer 1. This study aims to use the results from the CNN network convolution to aid researchers studying Thangka by helping to distinguish the outlines of the image effectively.

Keywords: Deep Learning; Thangka; CNN Network; line contours; art preservation

INTRODUCTION

Thangka is a popular art method in Tibet, China. It is often hand-drawn with brilliant colors on paper, silk, and clothes. Thangka has existed for hundreds of years and is an essential aspect of Tibetan culture, as it consists of many stories, folktales, and events related to their religion, Buddhism. "Its content involves interesting and colorful stories relating to historical events, religion, personage, local conditions and customs, folklore, fairy stories, building layouts, astronomy calendar, Tibetan medicine, Tibetan pharmacology, and so on." (1) Thangka are often hand drawn because of its size, and it often takes an extremely long time for one skillful artisan to finish a painting.

It is very popular in the Tibetan community, and many merchants and tourists purchase them to admire their beauty. Due to the uniqueness of the material, Thangka is often relatively small and easy to transport as it was throughout trade. However, it is usually scrolled up or folded, which eventually damages the artwork as the paper decays over a long period of time. Thus, preserving

Corresponding author: Xinyun Zhang, E-mail: 2048713048@qq.com.

Received August 13, 2024**; Accepted** September 21, 2024
https://doi.org/10.70251/HYJR2348.237585

and mending the artwork is extremely important in order to understand and learn about the complex Tibetan art. “Because it is influenced by history and religion, Thangka was listed as the first batch of state-level non-material cultural heritage in 2006. ” (2)

Deep learning is used in artificial intelligence, creating large neural network models capable of making accurate data-driven decisions. It is used in a large part of modern-day life, as many search engines, such as Facebook, Google, Baidu, and Microsoft, integrate deep learning into their features for image search and machine translation (4). The data processing factor of deep learning can also be used in fields such as agriculture by analyzing big agricultural data and new information and communication techniques (4). In this research, I utilized deep learning to extract the outlines of Thangka so that study experts could better examine the fractions as they may be damaged or unclear. Deep learning is a computational method that utilizes multiple processing layers to learn data representations at various levels of abstraction (3).

By utilizing deep learning, I hope to achieve successful line drawing generation of Thangka; Previous studies have demonstrated line drawing generation across various fields, including the generation of line drawings from photographs. The study's objective is to use CNN to produce images that capture a photograph's defined edges, contour lines, and texture lines (10). Further studies of the photo line drawing generation involve human portraits. While human portrait line drawing generation is wildly popular for its minimalistic aesthetics, it also poses various challenges regarding the line generation itself, as it often involves identifying valuable information within pixels. In order to address these challenges, researchers developed a new network named PLDGAN (Portrait Line Drawing Generative Adversarial Network), which enables clear photo line depiction (11).

Additionally, this study is closely related to edge detection, a technique in image recognition in image recognition that identifies sharp edges and discontinuities in images (12, 13). Edge detection can be divided into two categories: gradient-based and Laplacian-based. The main difference between the two categories is that: “The gradient method detects the edges by looking for the maximum and minimum in the first derivative of the image. The Laplacian method searches for zero crossings in the second derivative of the image to find edges “(14)” Edge detection involves the development of various operators such as the Sobel Operator, Robert’s cross Operator, Laplacian of Gaussian, and the Canny Edge Detection Algorithm. The issues in edge detection involve false line detection, missing edges, sensitivity, etc. Through testing, these operators show that the Canny Edge Detection Algorithm yields the best result. However, it is still essential to choose the most suitable operators for different projects (12).

Despite Thangka’s cultural and historical significance in Tibet culture and Buddasim, few studies have combined deep learning with Thangka preservation efforts. Thus, this study is dedicated to facilitating future research on Thangka by providing a foundation that leverages deep learning to support its preservation.

MATERIAL AND METHODS

I used different assorted images of Thangka from the internet; the data comes from these images. All the images are publicly accessible. I then included the images in Anaconda in a Python environment on a Surface Book Pro. The Anaconda is also equipped with PyTorch. Anaconda is a platform used for learning Python, data science, and machine learning while also assisting its users in building their projects. Python is a wildly popular coding language that is free and accessible for all users to code and study. Its versatile nature and increased productivity attract many coders from all levels.

In this research, I utilized convolutional neural networks (CNN). “In 1959, Hubel & Wiesel (1) found that cells in the animal visual cortex are responsible for detecting light in receptive fields. Inspired by this discovery, Kunihiko Fukushima proposed neocognition in 1980 (2), which could be regarded as the predecessor of CNN.” In deep learning, a convolutional neural network (CNN) is a class of artificial neural networks most commonly applied to analyze visual imagery. (3) “neural networks are a subset of machine learning. They are at the heart of deep learning algorithms. They are composed of node layers containing an input layer, one or more hidden layers, and an output layer. Each node connects to another and has an associated weight and threshold. If the output of any individual node is above the specified threshold value, that node is activated, sending data to the next layer of the network. Otherwise, no data is passed along to the next layer of the network. (4) ”

In CNN, there are the Input layer, hidden layers, and output layers. The convolution layer is a layer of image processing that exists in the hidden layer along with the pooling layer and fully-connected layer (15, 19). Figure 1 is a visual example of a convolution layer. The convolution layer involves a convolution kernel and parameter, input, etc. It extracts fundamental information from the input picture using kernel filters and then outputs a feature map (19). The kernel's size is smaller than that of the input volume, slides over the input step by step, and the kernel is dependent on the size input, although it is recommended that kernel size is odd so that there would be focus in the middle. The step can also be altered to reflect the desired results for the area where the kernel slides over the input values of all pixels (19, 17). The kernel also has values of its own, and that value times the value of those of the individual pixel in the input area, adding up to a new value (19). Additionally, to obtain the entire amount of information on the edge of the input picture, the convolution layer could fill in the additional information following the edge. In other words, the individual parameters go over the input value and calculate the dot product of the kernel pixel and the input pixels; the calculation is later stacked up into a feature map (16, 17).

After the convolution later outputs the feature map, the pooling layer begins to downsample the feature map. Although the feature map of the first convolution layer contained valuable information, the feature map size is still considerable due to the additional filling. The pooling layer is introduced to increase the calculation process's efficiency and gradually decrease the input dimension (21, 22). The pooling layer can be obtained by aggregating information from the previous feature map (22, 23, 24). The pooling layer can also control overfitting in the network (22, 24) and reduce the number of parameters and weights (24). The pooling layer divides the feature map into regions that do not overlap, named the pooling region. The pooling region's size can also be adjusted. One kind of pooling layer is max pooling, in which the pooling region picks out the maximum value of the region. This results in discarding trivial information and obtaining the defined and relevant details (24). Max pooling is optimal for the research purposes. Figure 2 is an example of max pooling where the greatest value from each color block is picked. While max pooling is the method that was utilized in this study, it is not the only type of pooling that exists (23, 24). Average pooling is another type of pooling method that, instead of selecting the maximum value like max pooling did, selects the average value of the pooling region, thereby yielding the average features presented in the patch instead of the most prominent ones in max pooling (23, 24).

Another method used to help the CNN network is Batch Normalization, also known as BN. The issue that exists in the process is that the distribution of each layer' input changes during training (36). Hence, the parameters also change along with the changes in the layer's input. Batch Normalization helps to accelerate the process of deep learning by standardizing the learning process and normalizing activations in layers of deep learning (35, 36).

In addition to convolution and pooling layers, the activation function plays a significant role in the CNN network. The activation function introduces non-linearity into the output value so that the networks can learn more progressively and become more versatile (25,26). Specifically, it decides which neuron to activate and which not (25). Although linear functions are easy to

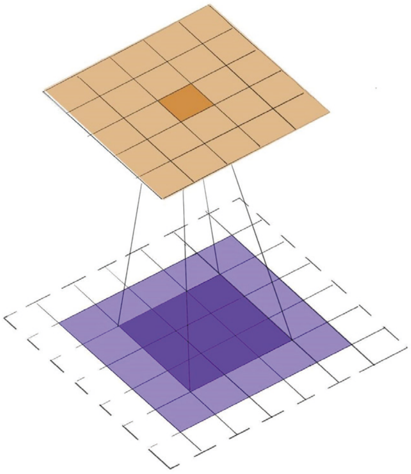

Figure 1. A visual example of a convolution layer in retrospect. The purple layer is the input layer. The darker purple square is the kernel which slides over the input layer to calculate and generates a new value. The new values form the feature map, which is the orange layer.

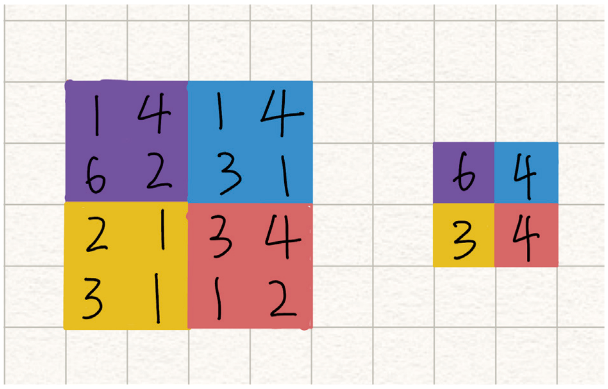

Figure 2. An example of max pooling where the maximum value from each region is picked and formed a new result.

train, they fail to learn complex mapping functions. Thus, using a nonlinear function is crucial for the learning and development of the neurons (25). There are various activation functions such as rectified linear (ReLU), Sigmoid, Tanh, etc. (25, 26, 27). A ReLU function is a piecewise nonlinear function presented below (25). ReLU function gives an output x if x is positive, 0 if otherwise (25), which leads to a faster operation. The ReLU function can learn at a very efficient rate and outshine the other activation function options. This Sigmoid function, on the other hand, resembles an S shape. The sigmoid function transforms the input value between 0.0 and 1.0 (27). Although the sigmoid function was wildly popular during the early 1990s regarding neural networks (27), the ReLU function has surpassed it. Figure 3 demonstrates how due to its limited sensitivity and the risk of gradient disappearing, the sigmoid function is obscured by the ReLU function (26, 27).

Network training involves choosing the proper loss function. Loss and gradient functions are interconnected; In order to get a minor loss, the functions must be changed accordingly. It is desired that a lesser error rate demonstrates that the model is well-trained. “The function we want to minimize or maximize is called the objective function or criterion. When we are minimizing it, we may also call it the cost function, loss function, or error function.” (37) One of the loss functions is cross entropy, in which the model minimizes the difference between the predicted value in the given data set and the value in the training data set. Another type of loss function also exists such as mean squared error function. Mean square error function is used for function approximation (regression problems), in the other cases, most neural networks contemporarily use cross entropy for their loss function.

Learning rate is another crucial aspect of Deep learning neural networks. It is involved in Stochastic gradient descent, which means it estimates the error gradient for the model's current state. The amount of weights updated with the training change is known as the learning rate. Learning rate is the most crucial hyper-parameter to the model's learning process as it determines how fast the model learns. It is also a very delicate balance between fast and slow learning. If it were to go too slow, the training process may take a significant amount of time and lose efficiency, whereas if it were to go too fast, it may miss the lowest error rate and lead to an eventual unstable training process.

ResNet stands for Residential Network, a network branch under the CNN network umbrella (31, 32). ResNet can be applied to numerous different fields, including medicine (29), weather classification (30), agriculture (33), etc. Before the development of ResNet, several attempts had been made to solve the issue of gradient disappearance (32). ResNet’s uniqueness is that it can skip one or more layers to resolve gradient disappearance (31). ResNet also includes many variants, such as ResNet 50, Resnet34, and Resnet 18. ResNet18, used in this study, signifies that the network has 18 layers (31, 32).

In the training process described in Figure 4, the decrease of the training loss means that the function is

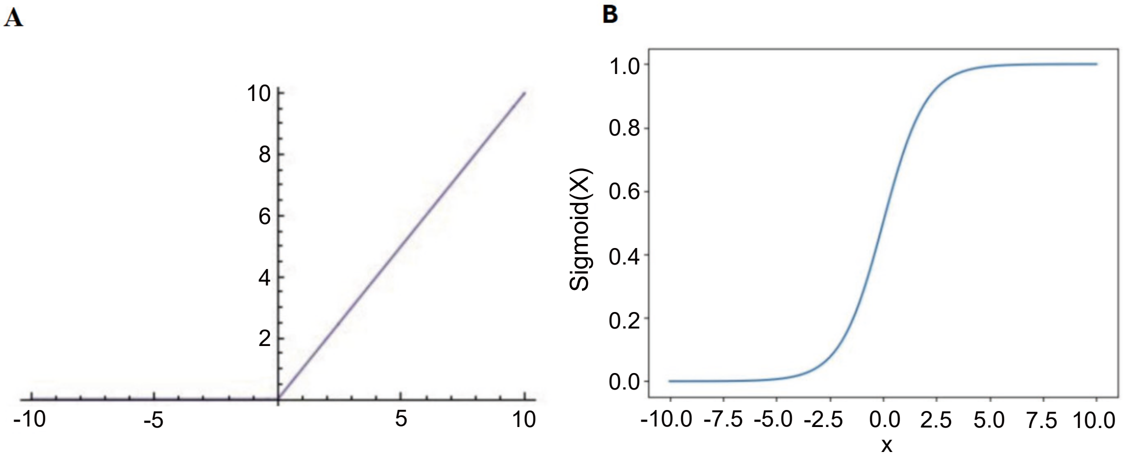

Figure 3. A shows the ReLu function, which has a linear compound to makes it easier to optimize; B shows the sigmoid function, which is nonlinear, meaning that any small changes in the x value may result in a dramatic change in the y value.

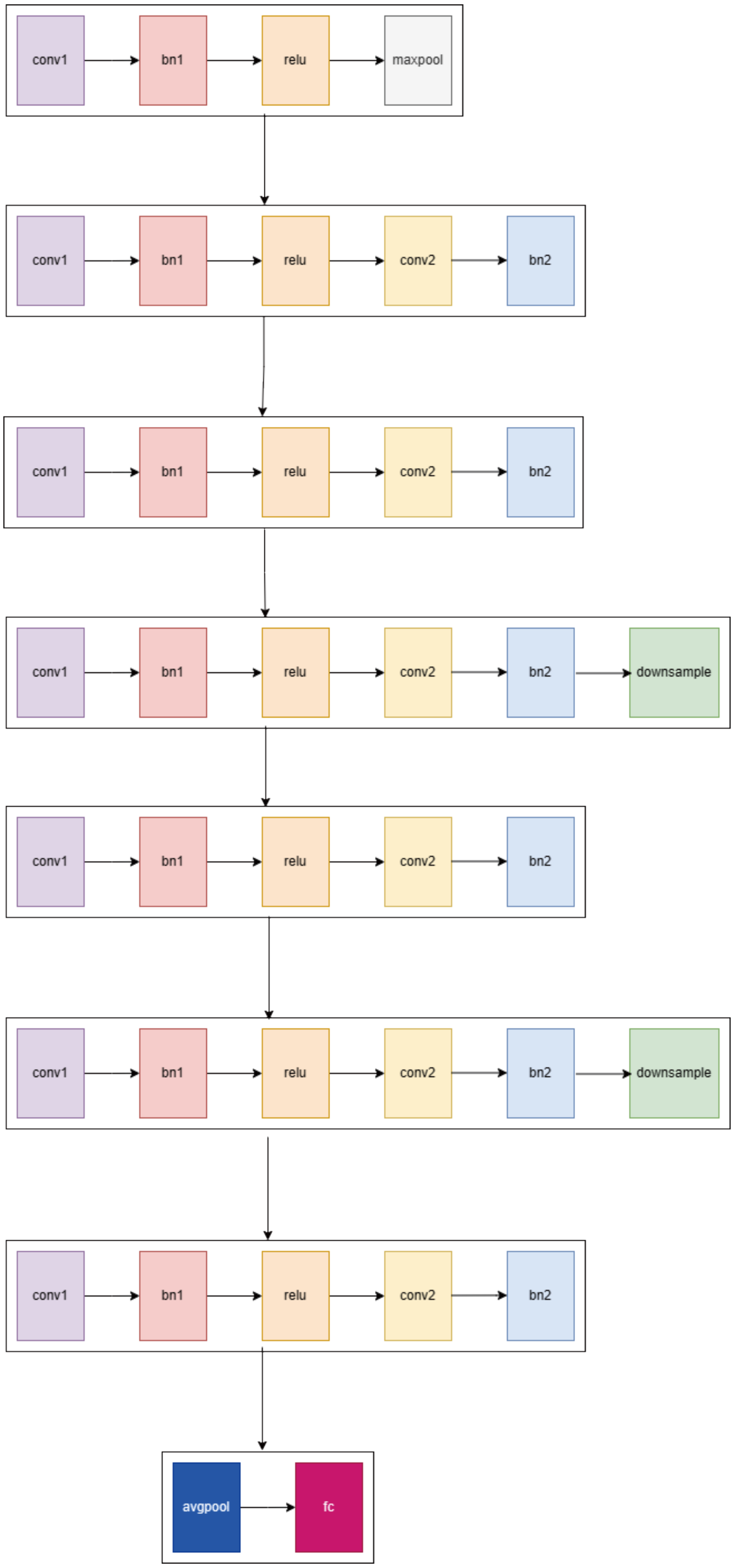

Figure 4. A diagram demonstrates how the image was convoluted in the study.

learning more accurately. After numerous training the training loss gradually approaches 0, which means that the function no longer needs to increase the rounds of training to make the function better. The training loss and the accuracy curve is nearly inversely disproportionate to each other. As shown in Figure 8, the more the training loss approaches 0, the more the accuracy curve approaches 1.

RESULTS

This is a picture of Thangka found on the internet.

As represented in Figure 5B, the edges of the outputs of conv1 are becoming more clear compared to the initial outputs. That is the evidence of going through the process of convolution and ReLU function and the development of

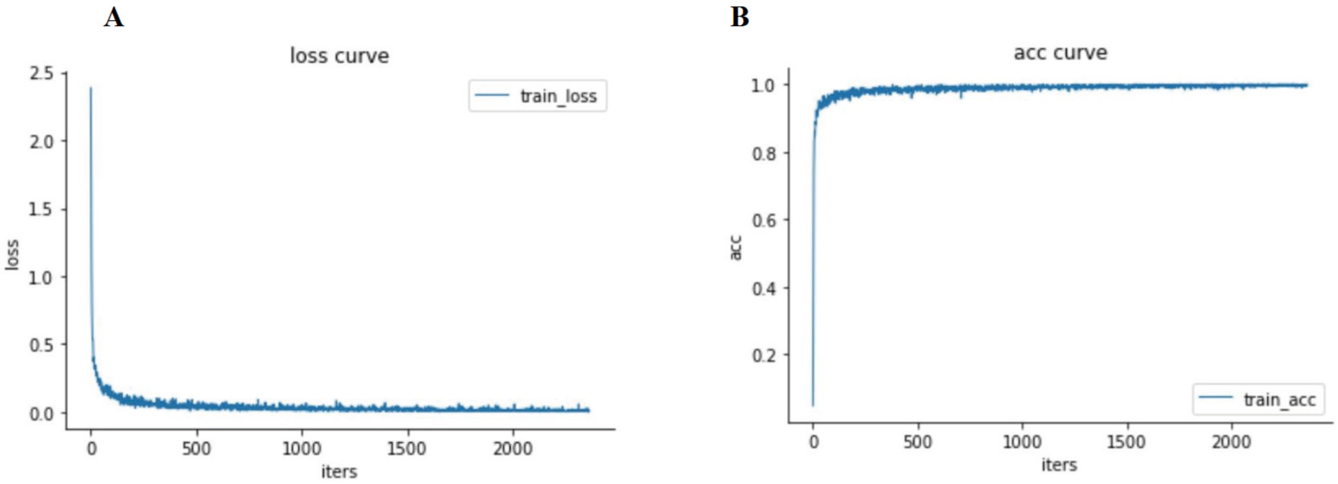

Figure 5. A shows the training loss approaches 0 after training; B is the accuracy curve.

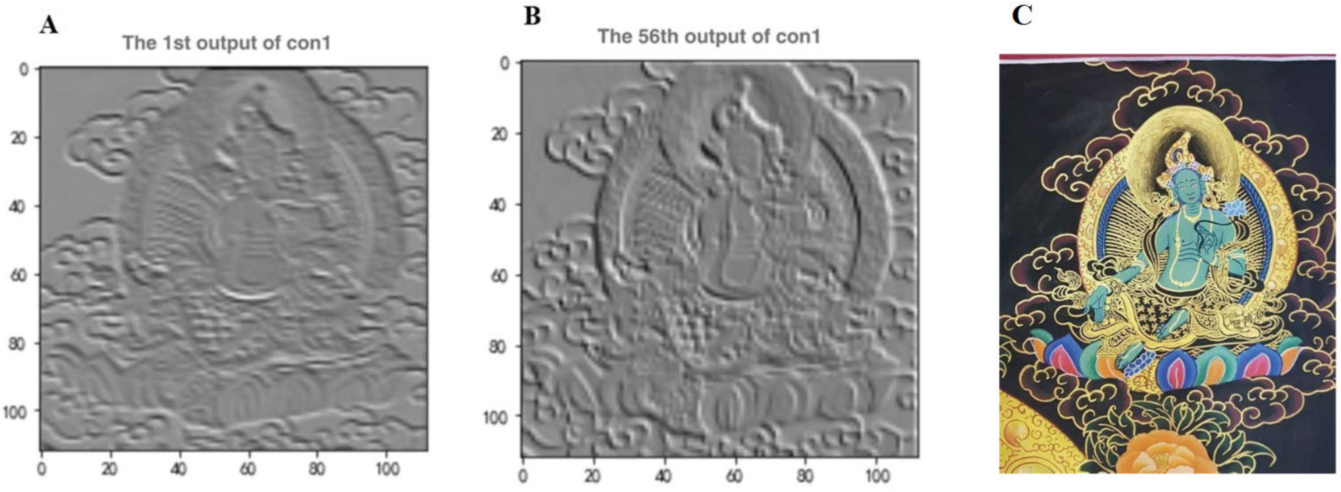

Figure 5. A, It indicates the first output of the convolution layer and the ReLU function; B, represents the 56th output of the conv1; C, demonstrates an image of a Thangka mural from the internet that was used in this experiment. This is the original image that was used to generate Figure 5A and 5B.

the function.

As it is shown in Figure 6B, the more the original photo of Thangka undergoes convolution and ReLU function, the more the defined edges are extracted and accentuated. In comparison of the outputs from conv1 and the outputs from layer 1, it is clear that the outputs in layer 1 are gradually becoming more abstract, as it is harder to recognize the original feature of the input picture. Some of the outputs may significantly differ from the other feature maps because the kernel size and weights are different.

From all the layers represented in Figure 7A, 7B and 7C, It is clear that through every cycle of convolution the output becomes more and more undistinguishable. At the 4th layer, the width and height of the output have diminished significantly. The 4th layer's largest width was 200, while the largest width at the 1st layer was 1600. This change in the width was because, through the process of convolution, pooling, and ReLU function, the size of

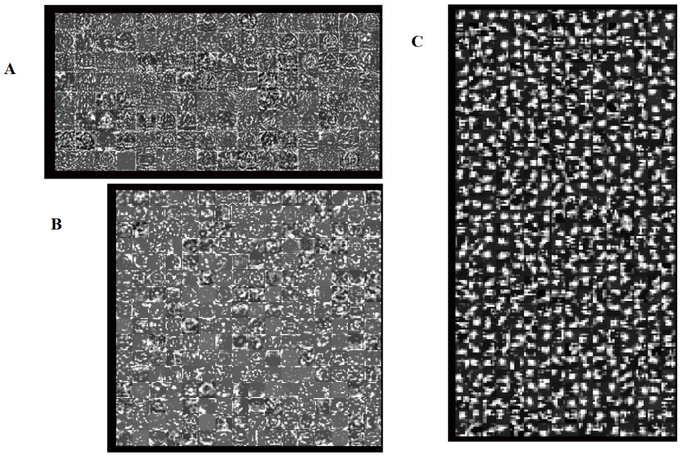

Figure 7. Shows all the outputs from the 2nd layer (A), all the outputs from the 3rd layer (B) and all the outputs from the 4th layer (C).

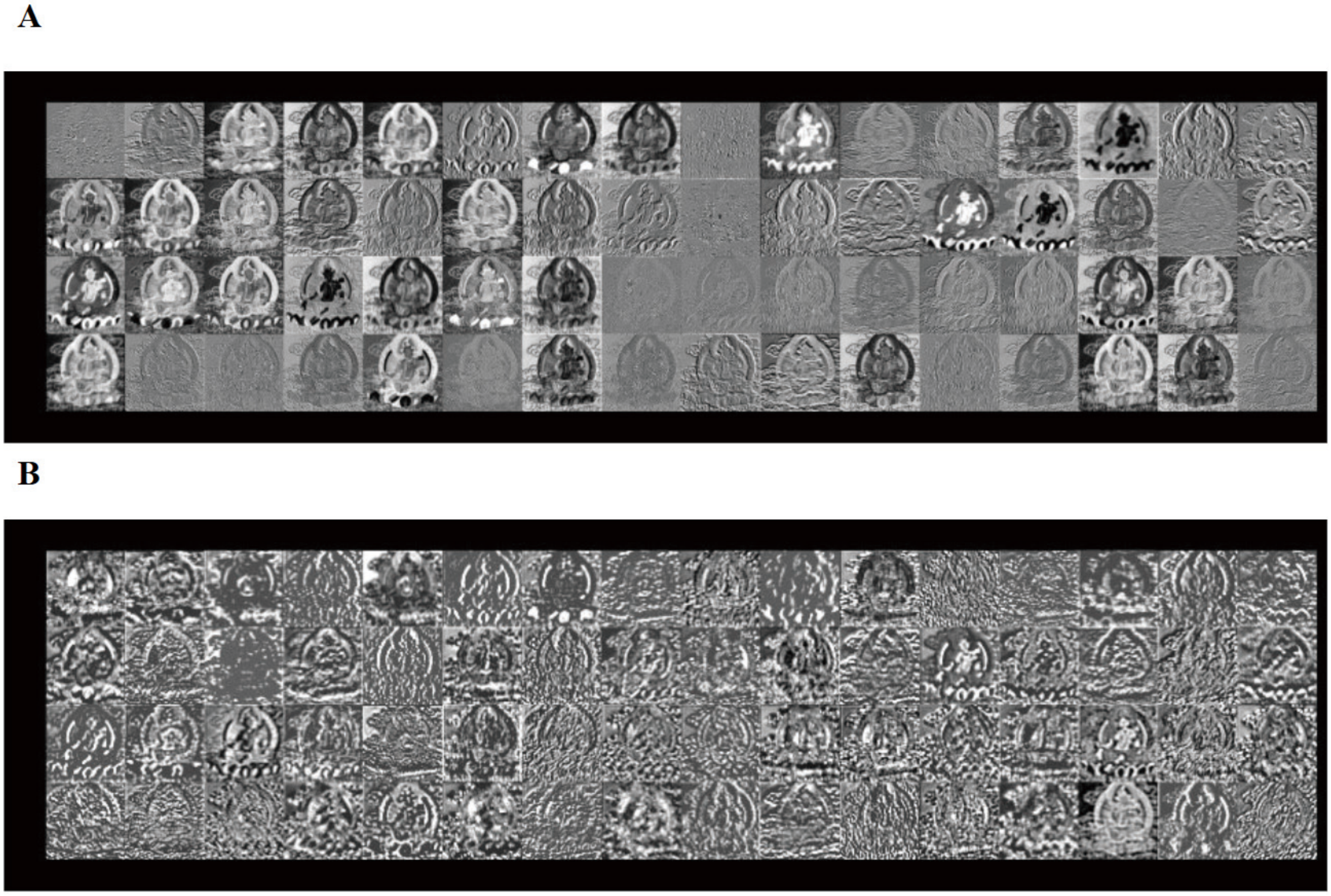

Figure 6. Shows all the outputs from conv1 (A) and the overview of all the outputs in layer 1 (B).

the feature map has become smaller. The content of the feature maps has also become more abstract, as seen in the 4th layer.

DISCUSSION

As the outputs of the Thangka reveal, the best results are generated during the first layer. With each convolution layer, the image becomes less ideal for future studies as it becomes more abstract.

However, even in the first layer, each output is different from the others as demonstrated in Figure 8. The two figures, the 48th and 3rd output, vary in background color, details, and line color. This variation is because they have different kernel sizes and weights. The clarity of the image, such as the 1st and 56th figures, is also very different for the exact cause. The ReLU function decides the neuron's decision of the feature map. This choice difference eventually led to the contrast in the final feature map. This representation is similar to carving out stamps. Some sculptors carve out the line on the image, while others carve out everything but the line. Even though the methods' eventual representation is opposite, both successfully produce a precise contour of the images. Thus, as long as the edge of the subject is clear, it is believed that feature map is successful in this study.

As shown in Figure 9, the 21st feature map is relatively similar to the 48th feature map; they both have the same

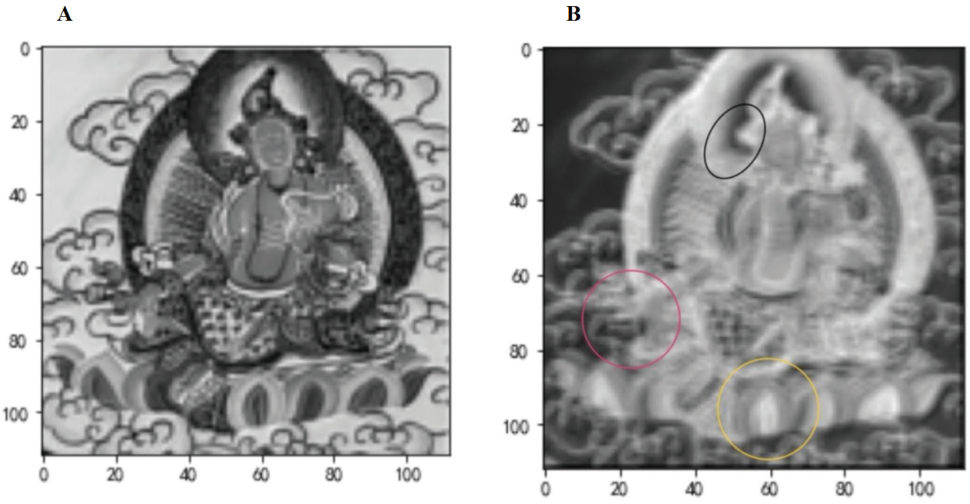

Figure 8. A represents the 3rd feature map; B represents the 48th feature maps from con1.

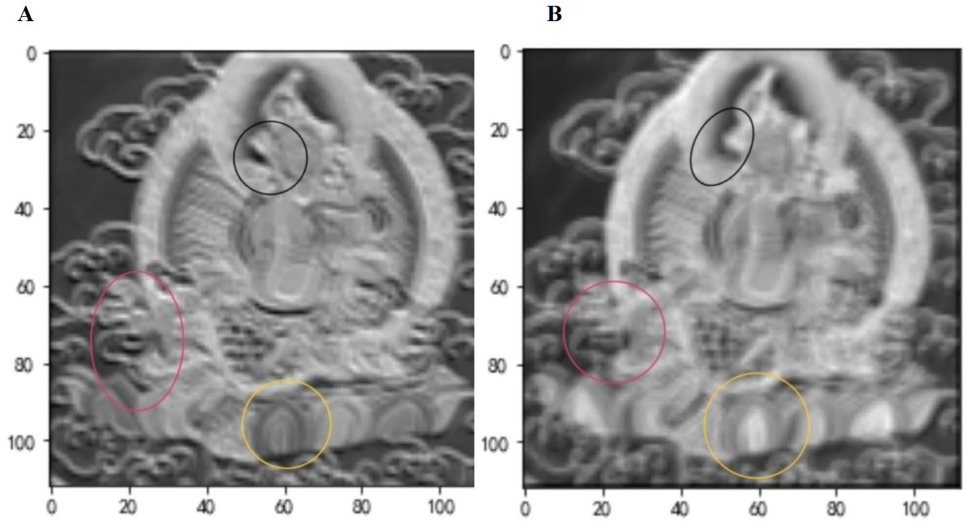

Figure 9. A shows the 21st feature map; B shows the 48th features maps from con1.

background color of black and the same white/gray lining and then it is between the 48th feature map and the 3rd feature map. However, they still represent some differences. For example, in the yellow circle, the pedal structure of the 48th feature map has a center that's light gray, while the pedal structure in the 21st structure is darker and more contrasted. Similarly, in the back circle of the 21st feature map, the structure of the painting is much darker, and the edge between the head of the deity and the halo is much more defined. The area In the pink circle resembles the same case, for the pink circle in the 21st feature map is also more contrasted and outlined, whereas the one in the 48th feature map is very blurry and unclear. The blurriness could be due to the clarity of the original input photo. Compared to the original image, the feature map also shows that the input has been cropped, and the orange flower at the bottom of the original photo beneath the deity has been cropped. Thus, it is vital for future studies to look through the generated feature maps and pick out which is the most ideal.

Other studies have also been conducted on similar objectives such as the study of Richer Convolutional Features for Edge Detection by NanKai University.

In the experiment, I used the same input picture as Nankai University and tested it in the model. Figure 10B shows the first feature map in the first con1. Both Figure 10B and Figure 10C showed more details than the results from Nankai University, such as facial features like the eyes, nose, and mouth. In the first feature map, it is visible to make out the eyes and the nose; in the results from Nankai University, it is difficult to distinguish the eye and the nose and other details on the face. Moreover, in the 11th feature map, it is even easier for more details to stand out, for example, the middle child's eyes and the polka dot pattern. Some of the hair strands of the middle child were also more visible. This phenomenon can be related to the learning process of the CNN network, which produces more detailed work as they learn.

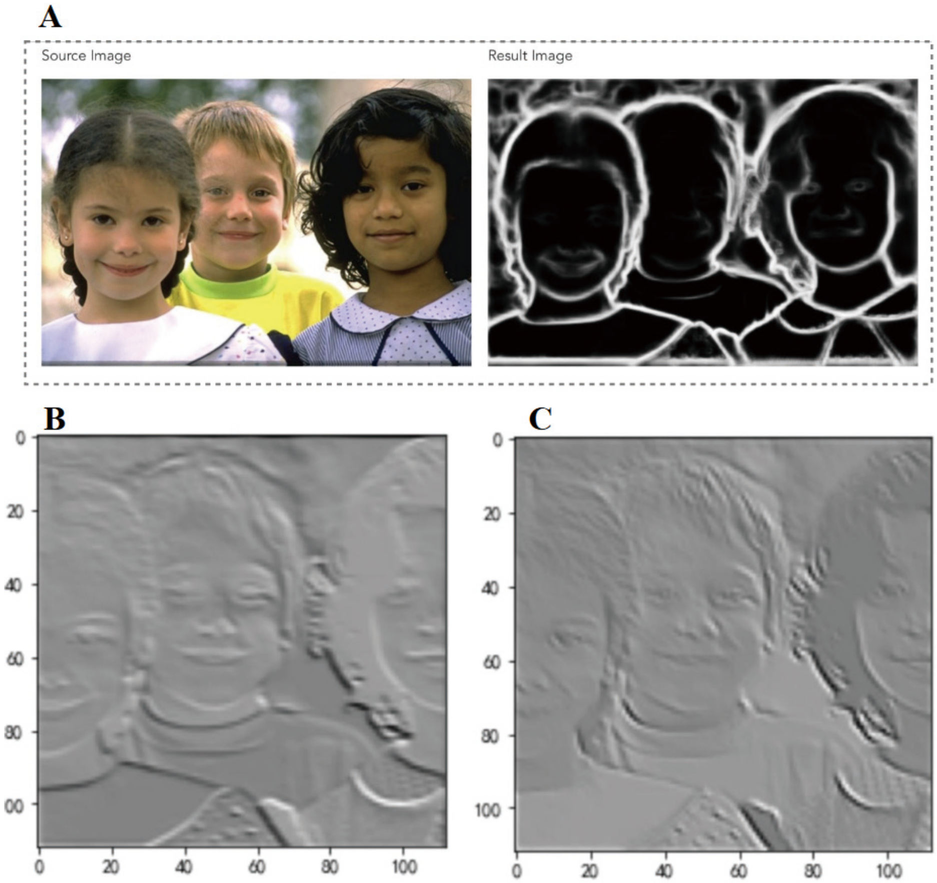

Figure 10. A reveals the original photo and the results from Nankai University; B shows the 1st output on con1 from the experiment; C shows the 11th output of con1 from the experiment.

CONCLUSION

This project looks into the utilization of deep learning in conserving Thangka culture, specifically in the Thangka drawings, which may have been damaged throughout the years after they were produced. Methods such as the CNN network, which has Resnet18, the specific mechanism, such as the convolution layer, pooling layer, and the ReLU function, etc. The results show that the feature maps produce successful line drawings of the contours of the Thangka. Results from other studies with the same purpose of outline generation are also compared as mine since they both yields similar results . Different studies have various methods of approaching the topic of generating line drawings. Still, the technique can also be more precise and contrasted in future studies that can be a worthwhile path to improve, such as using and developing better methods and models to help image recognition. Thangka is an essential cultural heritage for many scholars, researchers and historians.I developed this technique in hopes that the line contours could aid the understanding and documentation of the precious murals further and ultimately help preserve this valuable heritage so it can be shared with more people.

CONFLICT OF INTEREST

The author(s) declare that there are no conflicts of interest regarding the publication of this article.

REFERENCE

1. Yin Lu, Weilan Wang, and Danchun Yang. "Study on how to distinguish Thangka and non-Thangka image." International MultiConference of Engineers and Computer Scientists. 2010; 2.
2. Hu Wenjin, et al. "A new method of Thangka image inpainting quality assessment." Journal of Visual Communication and Image Representation. 2019; 59: 292-299.
3. LeCun, Yann, Yoshua Bengio, and Geoffrey Hinton. "Deep learning." nature. 2015; 521: 436-444.
4. Kelleher, John D. Deep learning. MIT press, 2019.
5. Convolutional neural network. Available from https://en.wikipedia.org/wiki/Convolutional_neural_network#cite_note-Valueva_Nagornov_Lyakhov_Valuev_2020_pp._232%E2%80%93243-1 (accessed on 2023-8-5).
6. What are convolutional neural networks ? Available from https://www.ibm.com/topics/convolutional-neural-networks (accessed on 2023-8-5).
7. Kamilaris, Andreas, Francesc X. Prenafeta-Boldú. "Deep learning in agriculture: A survey." Computers and electronics in agriculture. 2018; 147: 70-90.
8. Niu Xiao-Xiao, Ching Y. Suen. "A novel hybrid CNN–SVM classifier for recognizing handwritten digits." Pattern Recognition. 2012; 45 (4): 1318-1325.
9. Madjarov Gjorgji, et al. "An extensive experimental comparison of methods for multi-label learning." Pattern recognition. 2012; 45 (9): 3084-3104.
10. Inoue Naoto, et al. "Learning to trace: Expressive line drawing generation from photographs." Computer Graphics Forum. 2019; 38 (7).
11. Li Sifei, et al. "PLDGAN: portrait line drawing generation with prior knowledge and conditioning target." The Visual Computer. (2023): 1-12.
12. Shrivakshan GT, Chandramouli Chandrasekar. "A comparison of various edge detection techniques used in image processing." International Journal of Computer Science Issues (IJCSI). 2012; 9 (5): 269.
13. Ganesan P, G. Sajiv. "A comprehensive study of edge detection for image processing applications." 2017 international conference on innovations in information, embedded and communication systems (ICIIECS). IEEE. 2017.
14. Other Methods of Edge Detection. Available from https://www.owlnet.rice.edu/~elec539/Projects97/morphjrks/moredge.html#:~:text=However%2C%20the%20most%20may%20be,the%20image%20to%20find%20 edges. (accessed on 2023-8-5).
15. Dash Sujata, et al., eds. Deep learning techniques for biomedical and health informatics. Cham, Switzerland: Springer International Publishing. 2020.
16. Ke Qiuhong, et al. "Computer vision for human–machine interaction." Computer Vision for Assistive Healthcare. Academic Press. 2018; 127-145.
17. Khedgaonkar Roshni, Kavita Singh, Mukesh Raghuwanshi. "Local plastic surgery-based face recognition using convolutional neural networks." Demystifying Big Data, Machine Learning, and Deep Learning for Healthcare Analytics. Academic Press. 2021; 215-246.
18. Singh Sinam Ajitkumar, Takhellambam Gautam Meitei, Swanirbhar Majumder. "Short PCG classification based on deep learning." Deep learning techniques for biomedical and health informatics. Academic Press. 2020; 141-164.
19. Mostafa Sakib, Fang-Xiang Wu. "Diagnosis of autism spectrum disorder with convolutional autoencoder and structural MRI images." Neural engineering techniques for autism spectrum disorder. Academic Press. 2021; 23-38.
20. Convolutional Neural Networks (CNNs/ConvNets) https://cs231n.github.io/convolutional-networks/(accessedon2023-8-5).
21. Sun Manli, et al. "Learning pooling for convolutional neural network." Neurocomputing. 2017; 224: 96-104.
22. Gholamalinezhad Hossein, Hossein Khosravi. "Pooling methods in deep neural networks, a review." arXiv preprint

arXiv:2009.07485 (2020).

23. CNN | Introduction to Pooling Layer. Available from https://www.geeksforgeeks.org/cnn-introduction-to-pooling-layer/ (accessed on 2023-8-5).
24. Zafar Afia, et al. "A comparison of pooling methods for convolutional neural networks." Applied Sciences. 2022; 12 (17): 8643.
25. Activation functions in Neural Networks. Available from https://www.geeksforgeeks.org/activation-functions-neural-networks/ (accessed on 2023-8-5).
26. Banerjee Chaity, Tathagata Mukherjee, Eduardo Pasiliao Jr. "An empirical study on generalizations of the ReLU activation function." Proceedings of the 2019 ACM Southeast Conference. 2019.
27. A Gentle Introduction to the Rectified Linear Unit (ReLU). Available from https://machinelearningmastery.com/rectified-linear-activation-function-for-deep-learning-neural-networks/ (accessed on 2023-8-10)
28. Loss and Loss Functions for Training DeepLearning Neural Networks. Available from https://machinelearningmastery.com/loss-and-loss-functions-for-training-deep-learning-neural-networks/ (accessed on 2023-8-10).
29. Sarwinda Devvi, et al. "Deep learning in image classification using residual network (ResNet) variants for detection of colorectal cancer." Procedia Computer Science. 2021; 179: 423-431.
30. Al-Haija Qasem Abu, Mahmoud A. Smadi, Saleh Zein-Sabatto. "Multi-class weather classification using ResNet-18 CNN for autonomous IoT and CPS applications." 2020 International Conference on Computational Science and Computational Intelligence (CSCI). IEEE. 2020.
31. Deep Residual Networks (ResNet, ResNet -50) - 2-24 Guide. Available from https://viso.ai/deep-learning/resnet-residual-neural-network/ (accessed on 2024-6-17).
32. An Overview of ResNet Architecture and Its Variants. Available from https://builtin.com/artificial-intelligence/resnet-architecture (accessed on 2023-8-5).
33. Ramirez, Oscar Jhon Vera, José Emmanuel Cruz de la Cruz, Wilson Antony Mamani Machaca. "Agroindustrial plant for the classification of hass avocados in real-time with ResNet-18 architecture." 2021 5th international conference on robotics and automation sciences (ICRAS). IEEE. 2021.
34. Liu Yun, et al. "Richer convolutional features for edge detection." Proceedings of the IEEE conference on computer vision and pattern recognition. 2017.
35. Santurkar Shibani, et al. "How does batch normalization help optimization?." Advances in neural information processing systems. 2018; 31.
36. Ioffe Sergey, Christian Szegedy. "Batch normalization: Accelerating deep network training by reducing internal covariate shift." International conference on machine learning. pmlr. 2015.
37. Deeplearningbook-chapter9:ConvolutionalNetworks.(n.d.). https://www.deeplearningbook.org/contents/convnets.html.