

Sequence Prediction with CNNs: A Novel Machine Learning Approach to Morphology

Anonymous authors

Paper under double-blind review

Abstract

Morphological operations like dilation and erosion are foundational in binary image processing, yet their sequential application to transform one binary matrix into another remains an underexplored domain. This study investigates whether a convolutional neural network (CNN) can predict the sequence of transformations required to iteratively transform a binary input matrix A into a target output matrix B . The transformations are defined over a hypothesis class comprising 16 distinct operations, derived from applying dilation and erosion with eight unique structuring elements.

To evaluate this, we developed a synthetic dataset of binary matrix transformations, applying sequences of operations using predefined structuring elements, including cross, plus, rhombus, square, and directional line kernels. A CNN model was trained to classify the next transformation in a sequence, leveraging convolutional layers, batch normalization, and global average pooling for efficient spatial feature extraction. The model's performance was assessed using two metrics: next-in-sequence prediction accuracy and the ability to reconstruct full transformation sequences.

Experimental results demonstrate the model's robustness in generalizing across varying matrix sizes and sequence lengths. Notably, while the next-operation prediction accuracy was limited, the model achieved a remarkable 54.7% success rate in generating complete transformation sequences via alternative paths, often discovering shorter paths to achieve the target matrix. These findings reveal the potential of neural networks to approximate and optimize complex binary matrix transformations, opening avenues for further research into sequence learning and abstraction in image processing tasks.

1 Problem Statement

The objective of this project is to investigate the capacity of sequential morphological operations, such as dilation and erosion, to iteratively transform a binary input matrix A into a target output matrix B . These operations are applied using a set of predefined structuring elements, creating a hypothesis class of 16 distinct transformations. The transformation process can be mathematically represented as:

$$B = f_n \circ f_{n-1} \circ \cdots \circ f_1(A), \quad f_i \in S$$

where S is the set of transformations derived from applying dilation and erosion with eight unique structuring elements.

The project aims to evaluate the hypothesis that a neural network model can predict the sequence of transformations required to iteratively transform A into B . Specifically, the model is tasked with:

1. Predicting the next operation in the sequence at each step.
2. Reconstructing the full transformation sequence when A and B are given.

Performance will be assessed through:

1. **Next-in-sequence prediction accuracy:** The model's ability to correctly predict the next operation in a sequence.
2. **Full-sequence prediction accuracy:** The model's ability to accurately reconstruct the entire sequence of transformations leading to B .

The study investigates the influence of factors such as the choice of structuring elements, sequence length, and the complexity of the operations on model performance. Additionally, it evaluates the model's ability to generalize to unseen transformation sequences and its robustness when handling longer sequences. The ultimate goal is to determine whether the neural network can effectively learn and predict the transformation process across a diverse set of binary matrices and structuring elements.

2 Methodology

The objective of this project is to investigate the impact of image processing operations, particularly the morphological operations "dilation" and "erosion", on binary image sequences. The study focuses on generating a dataset of binary image transformations using different structuring elements (SE) and evaluating the ability of a Convolutional Neural Network (CNN) to predict the next transformation to be applied in sequence.

This study employs an experimental approach centered on sequential image processing. The primary morphological operations tested are dilation and erosion, applied using a predefined set of eight unique structuring elements (SE). The transformations are applied sequentially to a binary matrix A , producing an output matrix B after multiple steps. Each transformation in the sequence is drawn from a hypothesis class of 16 possible operations (8 dilations and 8 erosions).

A CNN model is trained to predict the next transformation in sequence (morphological operation and structuring element) applied to A to generate B . The architecture leverages convolutional layers to extract spatial features, followed by fully connected layers for classification. The model outputs one of 16 possible labels corresponding to the transformation to be applied next.

2.1 Dataset Generation

Synthetic datasets were created by performing sequential dilations and erosions on randomly generated binary matrices of fixed size $N \times N$. The dataset is generated as follows:

- Random binary matrices of size $N \times N$ serve as the initial input A .
- A sequence of morphological operations is applied iteratively to A using randomly selected structuring elements, resulting in a final output matrix B .
- Each sample in the dataset consists of an input-output pair (A, B) and the label corresponding to the last transformation applied in the sequence.

Since a generating function was used for the datasets, both training and testing data were generated separately using this function.

The structuring elements used in the dataset are as follows:

- **SE1 (Checkerboard Pattern):** A kernel with alternating squares, creating a checkerboard effect.
- **SE2 (Cross):** A cross-shaped kernel, focusing on the vertical and horizontal regions intersecting at the center.

- **SE3 (Inverted Checkerboard):** A reversed checkerboard pattern compared to SE1, with opposite color alternations.
- **SE4 (Hollow Square):** A square-shaped kernel with a hollow center, affecting only the border regions.
- **SE5 (Vertical Right Half):** A kernel that emphasizes the right half of the matrix, split vertically.
- **SE6 (Vertical Left Half):** A kernel that emphasizes the left half of the matrix, split vertically.
- **SE7 (Horizontal Top Half):** A kernel that emphasizes the top half of the matrix, split horizontally.
- **SE8 (Horizontal Bottom Half):** A kernel that emphasizes the bottom half of the matrix, split horizontally.

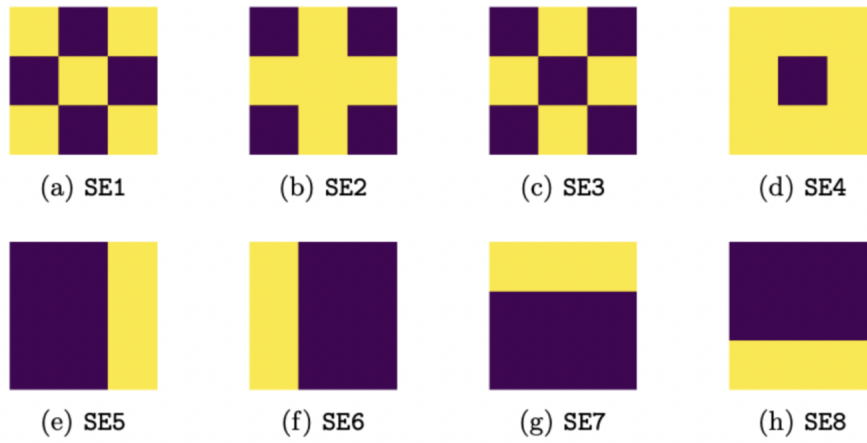


Figure 1: The structuring elements for the morphological operators are depicted above.

Each structuring element is applied in either dilation or erosion, which respectively expand or shrink the white regions of the binary image. The complexity of each sequence is controlled by varying the number of operations applied, allowing for both simple and complex transformation sequences to be tested.

2.2 Model Architecture

The proposed model is a Convolutional Neural Network (CNN) designed specifically to classify sequential morphological operations on binary matrices. It features a robust architecture that effectively balances spatial feature extraction, computational efficiency, and model generalization.

The architecture begins with an input layer tailored for paired binary matrices of size $N \times N$, represented as two channels (input matrix A and output matrix B). The core of the model consists of three convolutional layers, each equipped with ReLU activation and batch normalization to stabilize training and prevent issues like vanishing or exploding gradients. The convolutional layers use filters of size 3×3 with the same padding and He initialization, enhancing the model's ability to learn diverse patterns in binary matrices.

Key components of the architecture include:

1. **Feature Extraction:** The convolutional layers progressively extract spatial features, crucial for understanding the structural changes induced by morphological operations. Max pooling is employed after the first set of convolutions to downsample feature maps, reducing spatial dimensions and computational complexity while retaining the most significant features.

2. **Global Average Pooling:** Following the final convolutional layer, global average pooling condenses the spatial information into a single value per feature map. This operation enhances computational efficiency, reduces the risk of overfitting, and ensures the model focuses on global transformation patterns rather than localized noise.
3. **Fully Connected Layers:** The condensed features are passed through two dense layers with 64 and 32 neurons, respectively. These layers refine the feature representation, enabling the model to accurately classify the next operation in the sequence. The final dense layer outputs probabilities across 16 classes (representing the possible morphological operations), using a softmax activation function.

The model is trained using the Adam optimizer with a learning rate of 0.001, and the loss function is sparse categorical cross-entropy. A batch size of 32 and a validation split of 20% were used during training to ensure efficient learning and evaluation.

By leveraging techniques such as batch normalization and global average pooling, the model achieves a balance between computational efficiency and predictive accuracy. This architecture is well-suited for learning complex sequential patterns in binary matrices, making it highly effective for the task of predicting and reconstructing morphological transformation sequences.

3 Experimental Results and Validation

The model was trained on datasets consisting of fixed-size binary matrices for one set of training data, with each input paired with a corresponding label representing one of 16 possible operations. For training, a total of 10,000 input-output pairs were generated for each sequence length from 1 to the maximum sequence length, denoted as *MaxSeqLen*. This resulted in training sets containing $10,000 \times \text{MaxSeqLen}$ samples. Training parameters included 50 epochs of the CNN, a batch size of 32 and 20% validation split. The testing used 1000 samples.

The first form of testing was conducted under varying conditions to validate the model's robustness and generalization. Specifically:

1. **Fixed *MaxSeqLen*, Varying Matrix Sizes:** The model was tested on datasets with fixed *MaxSeqLen* for varying matrix sizes to observe prediction accuracy trends for different input sizes.
2. **Varying *MaxSeqLen*, Fixed Matrix Sizes:** Testing was also performed with variable *MaxSeqLen*, keeping matrix size fixed, to assess the effects of capping the number of operations, on the model and its results.

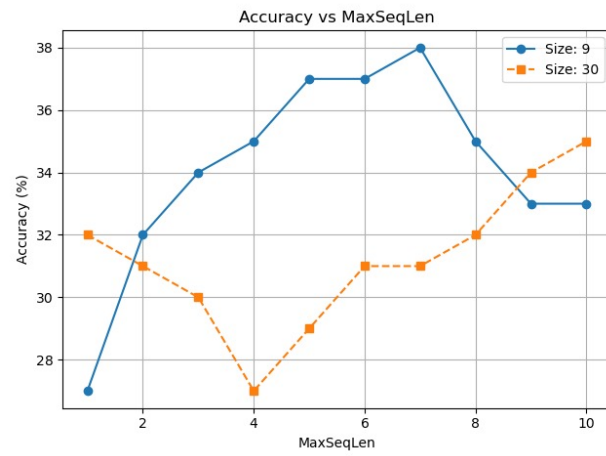
Here, a simple accuracy test of "prediction = true value" was used to find the percentage of correct predictions for the next operator in sequence. Trends were compiled in graphs for different conditions.

A second form of testing was to evaluate the usefulness of the "next - in - sequence" predictor to attempt to generate the whole sequence of operations from *A* to *B*. Here, a slightly different method of evaluation was used: starting at *A*, the model was allowed to predict the next operator in sequence and *A* was updated at each step, and taken as input for the next iteration, ignoring the actual value of the operator. The iteration halted if *B* was reached, indicating that the model was able to generate the output matrix from the input, or was terminated after a "maximum tries" condition was reached. Two parameters were noted: the percentage of matrices eventually reaching their output, and for the ones that did, the ratio of the sequence length the model took to the actual sequence length for the generated data.

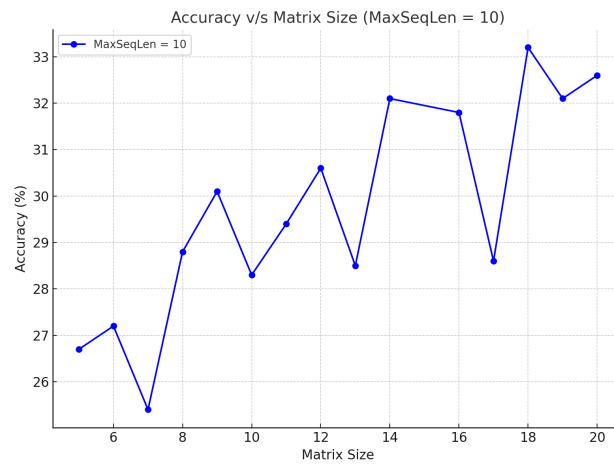
Two interesting results were found when the full sequence prediction was done on varying lengths:

- In spite of low correlation between the actual "next - in - sequence" operator and the prediction, upto 54.7% of inputs were able to reach the outputs via a **different path**.
- The ratio of average path size for the predictor versus actual was < 1 for longer sequences, often around 0.6, indicating that the model found shorter paths to reach the output.

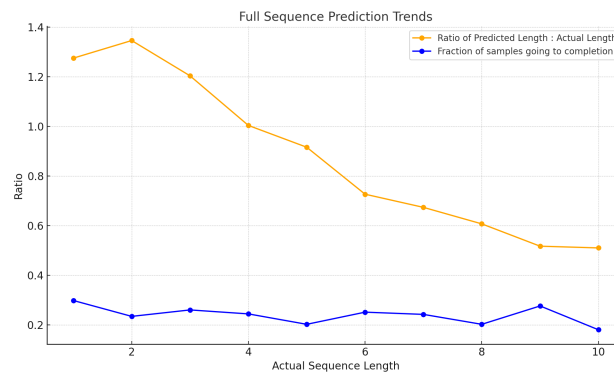
The following graphs indicate the trends observed in different data samples:



(a) Small and Large Matrices



(b) Matrix Sizes



(c) Full Sequence Trends

Figure 2: Combined figure showing various trends.

4 Conclusion and Future Work

This project explores the use of deep learning for classifying image transformations caused by morphological operations. Although the trained CNN model was able to identify some of the operators, the accuracy was around 25% for most cases, indicating the scope of improvement and experimentation with the model architecture and parameter tuning.

Key Findings:

- Accuracy of predictions reached a bottleneck for most datasets and sizes, though slightly higher accuracy was found for larger lengths and matrix sizes, possibly indicating that complex data is more dependent on the sequence and types of operations performed, whereas smaller data can be generated from a variety of paths.
- The full sequence prediction giving smaller sequences than the actual value indicates multiple, smaller routes to reach an output from an input.

Future Work:

- **Dataset Expansion:** Increasing the dataset size and including more complex images or operations could help fine tune the model and incorporate a variety of features.
- **Model Improvements:** Exploring other network architectures, hyperparameters, or optimization techniques may help enhance accuracy.
- **Generalization to Real-world Data:** Applying the model to real-world datasets, such as those in medical or industrial imaging, could help assess its practical applicability.

The exploration of structuring elements and their impact on image transformations highlights opportunities for advancing image analysis and computer vision. This study serves as a foundation for further experimentation and development, paving the way for more robust models and real-world applications.