Machine Learning Project Report

Arjun Krishnan (2022A4PS0617G)

1 Problem Statement

The objective of this project is to evaluate the capability of current Machine Learning (ML) algorithms in performing image transformations based on visual patterns, specifically using the IPARC dataset.

The hypothesis is: A fully connected neural network can accurately predict which operator was used in an image transformation defined by a simple morphological operation. The extent to which the ML model can achieve high accuracy in classifying input-output image pairs into any one of the eight different operations will determine the validity of this hypothesis.

2 Methodology

2.1 Data Preparation

- There is a dataset made separately for each of dilation and erosion.
- The dataset comprises multiple JSON files, each containing pairs of input and output images (15x15 pixels each) along with a transformation label.
- A task is defined by an operation (dilation or erosion) and an operator (SE1 - SE8).
- There are 10,000 image pairs in each task file. The total of 80,000 image pairs (per model) are shuffled and split into the train and test sets.
- Preprocessing steps include normalizing pixel values and converting the data into tensors suitable for deep learning models.

2.2 Model Architecture

- A neural network with fully connected layers:
  - Input Layer: Flattens the input (2x15x15) images into a single-dimensional array.
  - Hidden Layers: ReLU activation is used.
  - Output Layer: Outputs one of the eight operators used, using the softmax function.

2.3 Training Process

- The dataset was split into training (90%) and testing (10%) subsets.
- Loss Function: Negative Log-Likelihood Loss (NLLLoss).
- Optimizer: Adadelta optimizer with a learning rate scheduler.
- Evaluation Metrics: Accuracy and loss per transformation class.

2.4 Adadelta Optimizer

The Adadelta optimizer was chosen due to its advantages over traditional stochastic gradient descent (SGD):

- No need for manual learning rate tuning.
- Efficient memory usage.
- Faster convergence, especially in complex ML tasks.

2.5 Model Selection

Various configurations were tested by changing the number of layers, batch size, and learning rate. Results are summarized below:

Model	Layers	Batch Size	Learning Rate	ReLU
D1	2	32	0.5	Yes
D2	3	32	0.5	No
D3	4	64	1.0	Yes
D4	3	32	0.6	Yes
D5	2	128	0.6	Yes

Table 1: Model configurations and key hyperparameters for dilation.

3 Experimental Results

3.1 Model Performance for Dilation

Test accuracies for the best-performing model (D1) are shown below:

3.2 Error Analysis for Dilation

- Best performance observed for SE5, SE6, SE7, and SE8 (near-perfect accuracy).
- Worst performance for SE3 and SE2 (49.23% and 53.09%).
- Reducing the number of layers generally improved accuracy.

Structuring Element (SE)	Test Accuracy (D1)
SE1	72.57%
SE2	53.09%
SE3	49.23%
SE4	69.48%
SE5	99.80%
SE6	99.79%
SE7	100.00%
SE8	100.00%

Table 2: Test accuracy for each structuring element (SE) for dilation (model D1).

3.3 Model Performance for Erosion

Model	Layers	Batch Size	Learning Rate	ReLU
E1	2	128	0.5	Yes
E2	4	64	0.8	Yes
E3	6	64	0.8	Yes

Table 3: Model configurations and key hyperparameters for erosion.

4 Conclusion and Future Work

4.1 Conclusion

The hypothesis is FALSE. The neural network performed well on simpler transformations but struggled with complex patterns.

4.2 Future Work

- Improve computational resources to use lower learning rates, higher batch sizes, and more hidden layers.
- Introduce CNNs for enhanced feature extraction.
- Use transfer learning for improved generalization.
- Explore hybrid approaches combining symbolic reasoning with ML techniques.
- Expand the dataset with more diverse transformations.

5 References

Inspired by the MNIST classification task for handwritten digit recognition:
`https://www.kaggle.com/code/imdevskp/digits-mnist-classification-using-cnn`