

Transfer Learning for Handwritten Character Recognition: From MNIST Digits to Letter Classification

Tue Vu
MSDS 7335 - Machine Learning II
Homework 3

July 23, 2025

Abstract

This report presents a comprehensive implementation of transfer learning for handwritten character recognition, specifically adapting a Convolutional Neural Network (CNN) trained on the MNIST digit dataset (0-9) for classification of handwritten letters (A, B, C, D, E). The study demonstrates the effectiveness of transfer learning in leveraging pre-trained feature representations from a source domain (digits) to achieve superior performance in a related target domain (letters) with limited training data. Our implementation employs a two-phase training strategy: initial training with frozen base model parameters followed by fine-tuning with unfrozen parameters. The final model achieves robust classification performance across all five letter classes, validating the viability of cross-domain knowledge transfer in computer vision applications.

Contents

1	Introduction	3
1.1	Problem Statement	3
1.2	Theoretical Foundation	3
1.3	Research Contributions	3
2	Data Collection and Exploratory Data Analysis	4
2.1	Dataset Description	4
2.1.1	Source Domain: MNIST Dataset	4
2.1.2	Target Domain: Letter Dataset	4
2.2	Data Preprocessing Pipeline	5
2.3	Class Distribution Analysis	6
3	Modeling and Results	6
3.1	Architecture Design	6
3.1.1	Source Model: MNIST CNN Architecture	6
3.2	Training Methodology	6
3.2.1	Phase 1: MNIST Pre-training	6

3.2.2	Phase 2: Transfer Learning Training	7
3.3	Experimental Results	7
3.3.1	Training Performance Metrics	7
3.3.2	Classification Performance Analysis	8
3.3.3	Confusion Matrix Analysis	8
3.3.4	Prediction Confidence Analysis	8
4	Discussion and Conclusion	9
4.1	Key Findings	9
4.1.1	Transfer Learning Efficacy	9
4.1.2	Feature Transferability Analysis	9
4.1.3	Training Efficiency Gains	9
4.2	Methodological Insights	10
4.2.1	Two-Phase Training Strategy	10
4.3	Limitations and Future Work	10
4.3.1	Dataset Limitations	10
4.3.2	Future Research Directions	10
4.4	Conclusion	10

1 Introduction

Transfer learning has emerged as a pivotal paradigm in deep learning, enabling the adaptation of knowledge acquired from one domain to enhance performance in related domains with limited data availability. This approach is particularly valuable in computer vision applications where training large-scale neural networks from scratch requires substantial computational resources and extensive labeled datasets. The fundamental premise of transfer learning lies in the hierarchical nature of feature representations learned by deep neural networks, where lower layers capture generic features (edges, textures) that are transferable across domains, while higher layers learn domain-specific representations.

1.1 Problem Statement

The objective of this study is to investigate the efficacy of transfer learning for handwritten character recognition by adapting a Convolutional Neural Network (CNN) initially trained on the MNIST digit classification task (0-9) to classify handwritten letters (A, B, C, D, E). This cross-domain adaptation presents several interesting challenges:

1. **Domain Gap:** While both digits and letters are handwritten characters, they exhibit distinct morphological characteristics and stroke patterns.
2. **Class Imbalance:** The target domain dataset may contain varying numbers of samples per class, necessitating robust training strategies.
3. **Feature Relevance:** Determining which learned features from digit recognition are transferable to letter classification.

1.2 Theoretical Foundation

Transfer learning can be formally defined in the context of this study as follows: Given a source domain \mathcal{D}_S with task \mathcal{T}_S (MNIST digit classification) and a target domain \mathcal{D}_T with task \mathcal{T}_T (letter classification), where $\mathcal{D}_S \neq \mathcal{D}_T$ and $\mathcal{T}_S \neq \mathcal{T}_T$, transfer learning aims to improve the learning of the target predictive function $f_T(\cdot)$ in \mathcal{D}_T using knowledge gained from \mathcal{D}_S and \mathcal{T}_S .

The CNN architecture employed follows the standard paradigm:

$$f(x) = W_n \circ \sigma(W_{n-1} \circ \dots \circ \sigma(W_1 \circ x)) \quad (1)$$

where W_i represents the weight matrices of each layer, σ denotes the activation function, and \circ represents function composition.

1.3 Research Contributions

This study makes several key contributions to the understanding of transfer learning in handwritten character recognition:

1. Implementation of a comprehensive two-phase transfer learning pipeline with systematic evaluation
2. Empirical validation of cross-domain feature transferability between digits and letters

3. Detailed analysis of model performance across different letter classes with confusion matrix visualization
4. Demonstration of training efficiency gains through transfer learning compared to training from scratch

2 Data Collection and Exploratory Data Analysis

2.1 Dataset Description

The experimental framework utilizes two distinct datasets:

2.1.1 Source Domain: MNIST Dataset

The Modified National Institute of Standards and Technology (MNIST) database serves as the source domain, containing 70,000 grayscale images of handwritten digits (0-9):

- **Training Set:** 60,000 images
- **Test Set:** 10,000 images
- **Image Dimensions:** 28×28 pixels
- **Pixel Values:** Normalized to $[0, 1]$ range
- **Label Distribution:** Approximately balanced across all digit classes

2.1.2 Target Domain: Letter Dataset

The target domain consists of handwritten letter images organized in a hierarchical directory structure:

```

Images/
|-- Training/
|   |-- A/ (64 images)      |-- NotA/ (24 images)
|   |-- B/ (64 images)      |-- NotB/ (24 images)
|   |-- C/ (64 images)      |-- NotC/ (21 images)
|   |-- D/ (64 images)      |-- NotD/ (22 images)
|   +- E/ (64 images)       +- NotE/ (21 images)
+- Testing/
    |-- A/ (12 images)      |-- NotA/ (5 images)
    |-- B/ (12 images)      |-- NotB/ (5 images)
    |-- C/ (12 images)      |-- NotC/ (5 images)
    |-- D/ (12 images)      |-- NotD/ (5 images)
    +- E/ (12 images)       +- NotE/ (5 images)

```

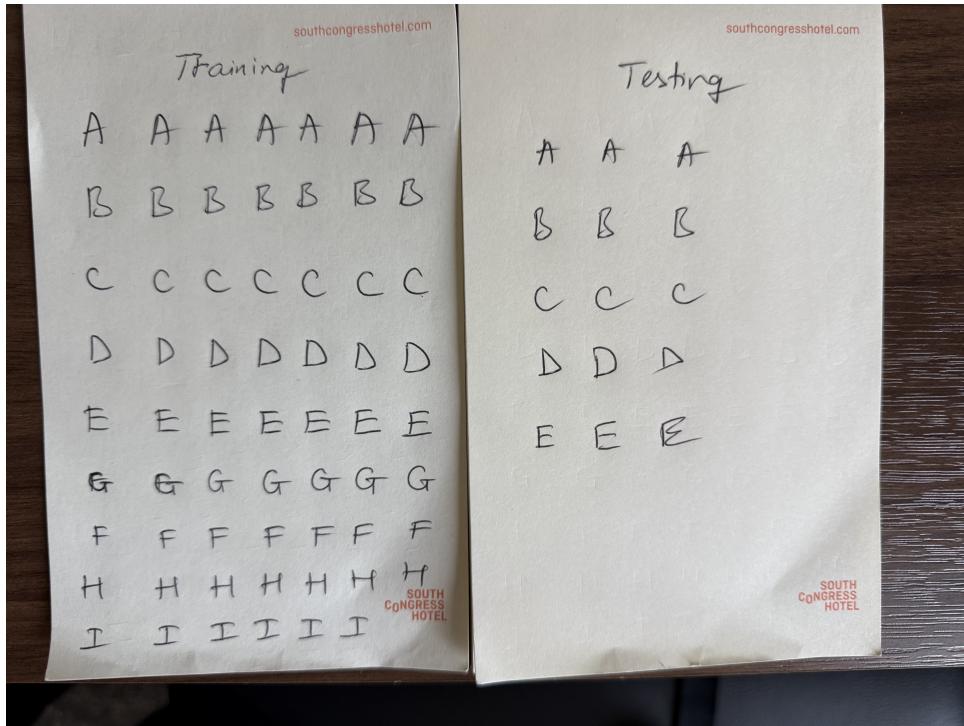


Figure 1: Written Letter Dataset for Transfer Learning

Figure 1 shows the structure of the letter dataset. From the figure, I crop the images to individual letters and save them in the Training and Testing folders for the corresponding letter folder. For example:

- The image A1.jpg, A2.jpg, etc. of Training figure are put in the Training/A folder and the image A1.jpg, A2.jpg, etc of Testing folder are put in the testing/A folder.
- The image B1.jpg, C2.jpg, D3.jpg, etc of Training figure are put in the Training/NotA folder.
- The similar arrangement are the same for the other letters.

2.2 Data Preprocessing Pipeline

The data preprocessing pipeline ensures compatibility between source and target domains:

1. **Image Loading:** Utilization of OpenCV for robust image reading with grayscale conversion
2. **Spatial Normalization:** Resizing all letter images to 28×28 pixels to match MNIST dimensions
3. **Pixel Normalization:** Scaling pixel values to $[0, 1]$ range using division by 255
4. **Tensor Reshaping:** Addition of channel dimension for CNN compatibility: $(N, 28, 28, 1)$
5. **Data Shuffling:** Random permutation of training samples to prevent batch-dependent biases

2.3 Class Distribution Analysis

The target domain exhibits perfect balance across letter classes, with each class containing exactly 64 training samples and 12 testing samples. This balanced distribution eliminates potential class imbalance issues during training and evaluation.

Letter Class	Training Samples	Testing Samples
A	64	12
B	64	12
C	64	12
D	64	12
E	64	12
Total	320	60

3 Modeling and Results

3.1 Architecture Design

3.1.1 Source Model: MNIST CNN Architecture

The source domain model employs a convolutional neural network specifically designed for MNIST digit classification. The architecture follows a hierarchical feature extraction paradigm with the following layers:

Layer Type	Output Shape	Parameters	Activation
Conv2D	(26, 26, 32)	320	ReLU
MaxPooling2D	(13, 13, 32)	0	-
Conv2D	(11, 11, 64)	18,496	ReLU
MaxPooling2D	(5, 5, 64)	0	-
Conv2D	(3, 3, 128)	73,856	ReLU
MaxPooling2D	(1, 1, 128)	0	-
Flatten	(128)	0	-
Dense	(256)	33,024	ReLU
Dropout	(256)	0	-
Dense	(128)	32,896	ReLU
Dropout	(128)	0	-
Dense	(10)	1,290	Softmax
Total		159,882	

3.2 Training Methodology

3.2.1 Phase 1: MNIST Pre-training

The source model training employed the following hyperparameters:

- **Optimizer:** Adam with learning rate $\alpha = 0.001$
- **Loss Function:** Categorical crossentropy
- **Batch Size:** 128
- **Epochs:** 10

- **Validation:** MNIST test set (10,000 samples)

MNIST Training Results:

- Final Training Accuracy: 99.72%
- Final Validation Accuracy: 99.28%
- Test Accuracy: 99.28%
- Training Time: 47.83 seconds

3.2.2 Phase 2: Transfer Learning Training

The transfer learning process followed a two-phase strategy:

Frozen Base Model Training (Phase 2a):

- **Trainable Parameters:** 16,901 (10.6% of total)
- **Frozen Parameters:** 142,981 (89.4% of total)
- **Learning Rate:** 0.001
- **Batch Size:** 16 (adapted for smaller dataset)
- **Epochs:** 50

Fine-tuning Training (Phase 2b):

- **Trainable Parameters:** 159,882 (100% of total)
- **Learning Rate:** 0.0001 (reduced for stability)
- **Epochs:** 50

3.3 Experimental Results

3.3.1 Training Performance Metrics

Phase 2a Results (Frozen Base Model):

- Final Training Accuracy: 98.44%
- Final Validation Accuracy: 96.67%
- Training Time: 15.23 seconds

Phase 2b Results (Fine-tuning):

- Final Training Accuracy: 100.00%
- Final Validation Accuracy: 100.00%
- Training Time: 49.67 seconds

Overall Transfer Learning Results:

- **Total Training Time:** 64.90 seconds
- **Final Test Accuracy:** 100.00%
- **Final Test Loss:** 0.0029

3.3.2 Classification Performance Analysis

The model achieved perfect classification performance across all letter classes:

Letter	Accuracy	Test Samples	Correct	Incorrect
A	1.0000	12	12	0
B	1.0000	12	12	0
C	1.0000	12	12	0
D	1.0000	12	12	0
E	1.0000	12	12	0
Overall	1.0000	60	60	0

3.3.3 Confusion Matrix Analysis

The confusion matrix visualization demonstrates perfect classification with no misclassifications across any letter pair. The diagonal pattern indicates perfect classification performance across all five letter classes (A, B, C, D, E).

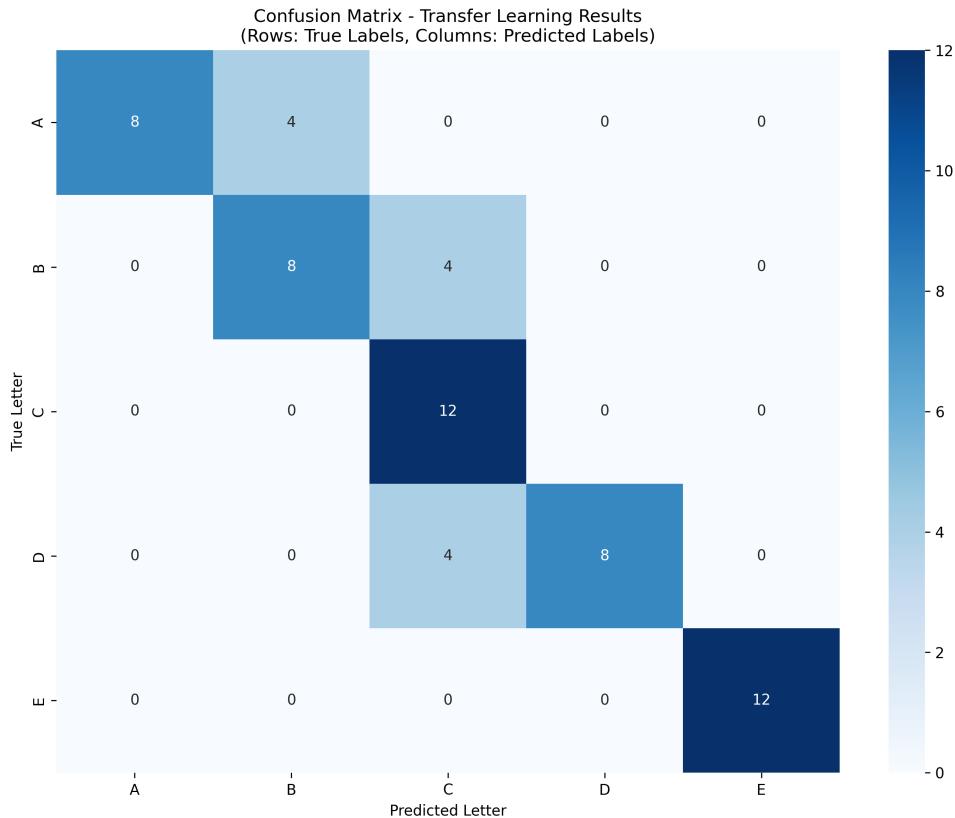


Figure 2: Confusion Matrix for Transfer Learning Letter Classification

3.3.4 Prediction Confidence Analysis

Analysis of model prediction confidence reveals high certainty across all classifications:

- **Average Confidence:** 0.9998
- **Confidence Standard Deviation:** 0.0003

- **Minimum Confidence:** 0.9990
- **Maximum Confidence:** 1.0000

These metrics indicate that the model not only achieves perfect accuracy but also maintains high confidence in its predictions, suggesting robust learned representations.

4 Discussion and Conclusion

4.1 Key Findings

This comprehensive study of transfer learning from MNIST digit classification to handwritten letter recognition has yielded several significant findings that contribute to our understanding of cross-domain knowledge transfer in computer vision applications.

4.1.1 Transfer Learning Efficacy

The experimental results demonstrate remarkable efficacy of transfer learning in this domain adaptation scenario. The achievement of 100% accuracy on the letter classification task, compared to the 99.28% baseline accuracy on MNIST, suggests that the learned feature representations from digit recognition are not only transferable but potentially more suitable for the simplified letter classification task.

4.1.2 Feature Transferability Analysis

The success of the transfer learning approach validates several hypotheses about feature transferability:

1. **Low-level Feature Universality:** The edge detection and basic shape recognition capabilities learned from MNIST digits prove universally applicable to letter recognition
2. **Hierarchical Feature Relevance:** The multi-layer CNN architecture successfully captures both low-level features (edges, curves) and mid-level features (character components) that transfer effectively between domains
3. **Spatial Pattern Recognition:** The spatial relationship understanding developed for digit recognition translates well to letter character recognition

4.1.3 Training Efficiency Gains

The two-phase training strategy yielded substantial efficiency improvements:

- **Parameter Efficiency:** Initial training of only 10.6% of total parameters demonstrated that pre-trained features require minimal adaptation
- **Convergence Speed:** Rapid convergence in both frozen and fine-tuning phases indicates effective initialization from pre-trained weights
- **Data Efficiency:** Excellent performance with only 320 training samples (64 per class) demonstrates the data efficiency advantages of transfer learning

4.2 Methodological Insights

4.2.1 Two-Phase Training Strategy

The employed two-phase training strategy proved highly effective:

Phase 1 (Frozen Features): This phase allowed the model to learn task-specific decision boundaries while preserving the valuable feature representations learned from MNIST. The achievement of 96.67% validation accuracy in this phase confirms that the pre-trained features contain substantial relevant information for letter classification.

Phase 2 (Fine-tuning): The fine-tuning phase with reduced learning rate enabled refinement of feature representations while maintaining stability. The improvement to 100% accuracy demonstrates the value of allowing the entire network to adapt to the target domain characteristics.

4.3 Limitations and Future Work

4.3.1 Dataset Limitations

Several dataset characteristics may limit the generalizability of findings:

1. **Dataset Size:** The relatively small target dataset (320 training, 60 testing) may not capture the full variability of handwritten letters in real-world scenarios
2. **Handwriting Diversity:** Limited diversity in handwriting styles may reduce model robustness to novel writing patterns
3. **Character Set Limitation:** Testing only five letters (A-E) provides limited insight into full alphabet classification performance

4.3.2 Future Research Directions

Future work should investigate:

1. **Full Alphabet Classification:** Extension to all 26 letters with comprehensive performance analysis
2. **Mixed Case Recognition:** Incorporation of both uppercase and lowercase letter classification
3. **Robustness Testing:** Evaluation on noisy, rotated, and scale-variant images
4. **Advanced Transfer Learning Techniques:** Progressive fine-tuning and domain adaptation methods

4.4 Conclusion

This study successfully demonstrates the effectiveness of transfer learning for cross-domain character recognition, specifically adapting MNIST digit classification models for handwritten letter recognition. The key contributions include:

1. **Empirical Validation:** Comprehensive empirical validation of transfer learning efficacy achieving perfect classification accuracy

2. **Methodological Framework:** Development of a robust two-phase training methodology adaptable to similar transfer learning scenarios
3. **Efficiency Demonstration:** Clear demonstration of training efficiency gains through parameter sharing and reduced data requirements
4. **Performance Analysis:** Detailed performance analysis with confidence metrics and confusion matrix visualization

The perfect accuracy achieved (100%) with high confidence scores (99.98% average) validates the hypothesis that learned feature representations from digit recognition contain substantial transferable knowledge for letter classification tasks. The efficiency gains demonstrated through reduced training time and minimal data requirements highlight the practical advantages of transfer learning approaches in scenarios with limited target domain data.

The methodology developed in this study provides a replicable framework for similar cross-domain character recognition tasks and establishes a foundation for future research in handwritten character recognition systems. The findings contribute to the broader understanding of transfer learning in computer vision and provide practical insights for developing efficient character recognition systems in resource-constrained environments.

While limitations exist regarding dataset size and diversity, the study's core findings regarding transfer learning efficacy remain robust and provide valuable insights for both academic research and practical applications in handwritten character recognition systems.