IMPACT OF DATA TRANSFORMATIONS ON MODEL ACCURACY: HOMOGRAPHY AND GAUSSIAN NOISE

Minwon Lee
Department of Automotive Engineering
Hanyang University
Seoul, South Korea

DECEMBER 6, 2024

Abstract

This report evaluates the robustness of various machine learning models, including Logistic Regression, SVM, MLP, and CNN, under two types of data transformations: Homography and Gaussian Noise. In autonomous driving perception systems, such transformations are the most common distortions encountered when recognizing numbers in road images. The results highlight model performance differences considering accuracy and suggest considerations for robust machine learning applications.

1 Problem Setting

The problem at hand is to evaluate the robustness of several machine learning models when faced with data transformations such as Homography and Gaussian Noise. The objective is to test how well models like Logistic Regression, FDA, SVM, MLP, and CNN perform under these challenging conditions, and to analyze the trade-offs between different models' ability to handle such transformations.

The models will be tested on the MNIST dataset. Before applying any transformations, the **baseline performance** of each model will be evaluated on the original (untouched) data to establish a reference point. This will allow us to compare how much the transformations affect the performance of each model.

2 Dataset and Data Split

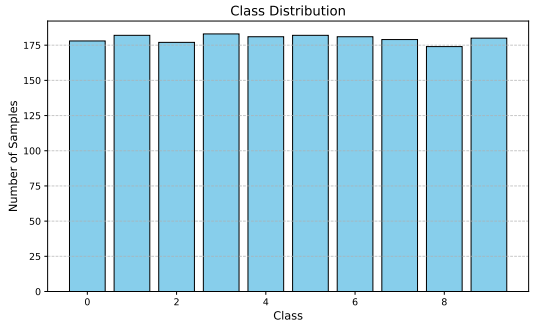
2.1 Dataset Overview

The dataset used for this study contains 1,797 samples, each represented by 64 features. This dataset consists of numerical data derived from handwritten digits. The dataset is uniformly distributed across 10 unique classes, representing digits from 0 to 9. This uniform distribution ensures that each class has roughly the same number of samples, which contributes to a balanced training process.

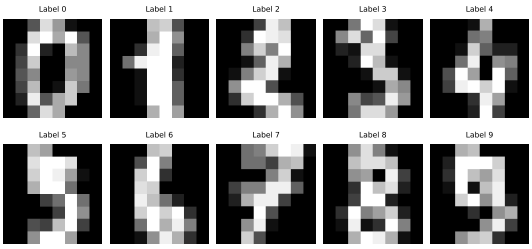
The dataset is split into training and test sets, with 1,437 samples used for training and 360 samples used for testing. This split ensures that the model can be evaluated on data it has not seen during the training phase.

2.2 Data Split

The dataset is split into training and test sets. The training set contains 1,437 samples, while the test set consists of 360 samples. The test set is reserved for evaluating the generalization performance of the trained model.



(a) Class Distribution Visualization



(b) Sample MNIST Images

Figure 1: Dataset Visualizations: Class distribution and example MNIST images.

2.2.1 Train and Validation Split with K-Fold Cross-Validation

To evaluate the robustness of the model and avoid overfitting, we employ 5-fold cross-validation on the training set. In 5-fold cross-validation, the training data is divided into 5 subsets (folds). The model is trained on 4 folds and validated on the remaining fold, and this process is repeated 5 times, with each fold serving as the validation set once.

The details of the data split for each fold are as follows:

- **Fold 1**: Training indices: 1,149; Validation indices: 288
- **Fold 2**: Training indices: 1,149; Validation indices: 288
- **Fold 3**: Training indices: 1,150; Validation indices: 287
- **Fold 4**: Training indices: 1,150; Validation indices: 287
- **Fold 5**: Training indices: 1,150; Validation indices: 287

3 Models and Validation Accuracy

3.1 Model Overview

The following machine learning models are used in this study to evaluate their robustness against data transformations:

- **Logistic Regression**: A linear model used for binary and multi-class classification, employing the logistic function to predict class probabilities.
- **Fisher Discriminant Analysis (FDA)**: A linear discriminant analysis model used to project data in such a way that maximizes class separability.
- **Support Vector Machine (SVM)**: Two versions are used in this study - one with an RBF kernel for non-linear classification and another with a polynomial kernel.
- **Multilayer Perceptron (MLP)**: A simple neural network model with a hidden layer, capable of learning complex non-linear patterns.
- **Convolutional Neural Network (CNN)**: A more advanced deep learning model particularly well-suited for image data.

3.2 Validation Accuracy for Each Model

The table below summarizes the validation accuracy of each model after performing 5-fold cross-validation on the training set. The values represent the average accuracy across the 5 folds.

The 5-fold cross-validation ensures that each data point in the training set is used for both training and validation, which helps provide a more reliable estimate of the model's performance and minimizes bias.

Table 1: Average Validation Accuracy for Each Model

Model	Average Validation Accuracy
Logistic Regression	0.9617
Fisher Discriminant Analysis (FDA)	0.9478
SVM (RBF Kernel)	0.9861
SVM (Polynomial Kernel)	0.9882
Multilayer Perceptron (MLP)	0.9763
Convolutional Neural Network (CNN)	0.9792

3.3 Preprocessing and Feature Extraction

Before training the models, the features are normalized to improve convergence during training. The feature values, originally in the range from 0.0 to 16.0, are scaled to a range of [0, 1] by dividing each value by 16.0. This normalization ensures that all features contribute equally during the training of the model, avoiding dominance by features with larger numerical values.

4 Experimental Setup

4.1 Homography Transformation

The homography transformation involves altering the perspective of the images to simulate distortions that might occur in real-world scenarios, such as capturing an image at an angle. In this study, we applied a perspective transformation by modifying the coordinates of the four corners of the image. Specifically, each corner's position was adjusted by adding a random offset within a maximum range specified by **max_offset**.

Three different levels of homography transformations were tested:

- **offset = 1.5**: Each corner was shifted by a random value within the range of [-1.5, 1.5] pixels.
- **offset = 2.0**: Each corner was shifted by a random value within the range of [-2.0, 2.0] pixels.
- **offset = 2.5**: Each corner was shifted by a random value within the range of [-2.5, 2.5] pixels.

This transformation was intended to test how well each model can handle changes in perspective. The range of offsets was chosen to progressively increase the severity of the transformation, thereby evaluating the robustness of the models under increasingly difficult conditions.

4.2 Gaussian Noise

Gaussian noise was added to the images to simulate noisy conditions that are often encountered in real-world image data, such as low-light conditions or sensor noise. Gaussian noise was introduced to the dataset at three different levels, each with a different value of standard deviation (**sigma**):

- **sigma = 0.2**: A moderate level of Gaussian noise.
- **sigma = 0.5**: An increased level of noise, introducing noticeable degradation in the image quality.
- **sigma = 0.7**: A high level of noise, significantly affecting image clarity and making the classification task more challenging.

The objective of introducing Gaussian noise was to evaluate the models' tolerance to noisy data. This is especially important in real-world applications where images are rarely pristine, and the robustness of the model under such conditions is crucial.

4.3 Transformation Visualization

To provide a qualitative insight into the applied transformations, Figure 2 illustrates the impact of Homography (max_offset = 2.5) and Gaussian Noise (sigma = 0.7) on sample images from the dataset. These examples represent the most challenging transformations applied in this study.

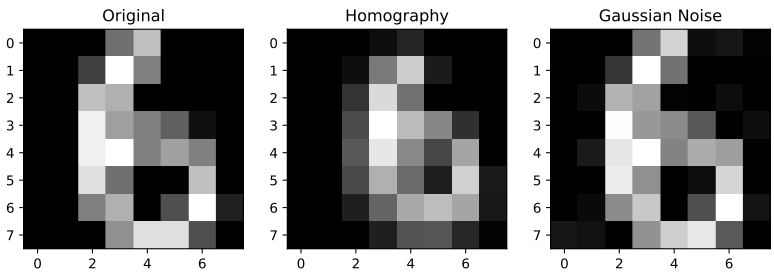


Figure 2: Visualization of Homography Transformation (max_offset = 2.5) and Gaussian Noise (sigma = 0.7).

4.4 Model Training and Evaluation

The models were trained on the original (untouched) training dataset and subsequently evaluated on:

1. The original test dataset (Baseline performance).
2. The homography-transformed datasets with max_offset values of 1.5, 2.0, and 2.5.
3. The Gaussian noise-transformed datasets with sigma values of 0.2, 0.5, and 0.7.

Each transformation type was tested at different severity levels, as described above, to understand how the model accuracy changes as the degree of transformation increases. This allowed us to analyze the strengths and weaknesses of each model when subjected to distortions common in autonomous driving environments, such as perspective shifts and noise interference.

5 Experiment Results

5.1 Baseline Accuracy

The baseline accuracy represents the performance of each model on the original MNIST dataset without any transformations. Figure 3 illustrates the baseline accuracy for all models, highlighting the effectiveness of CNN and SVM with a polynomial kernel, which achieved the highest accuracy.

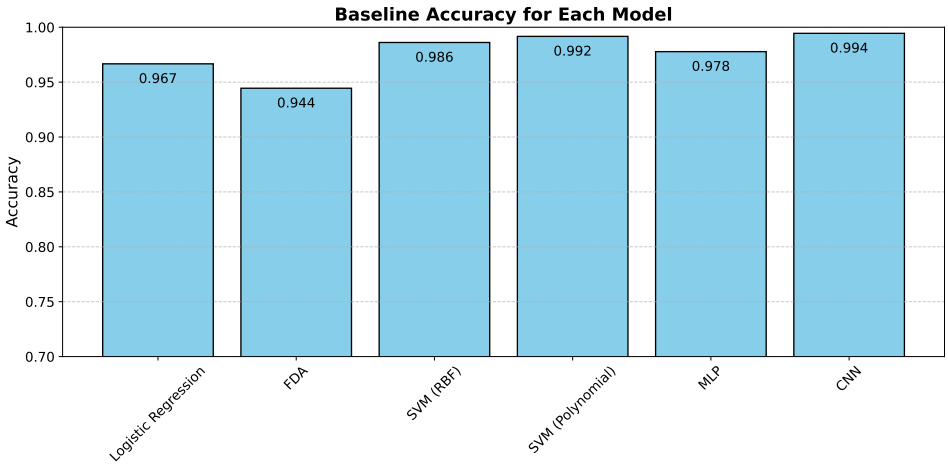


Figure 3: Baseline accuracy for each model on the original MNIST dataset.

5.2 Impact of Homography Transformation

The robustness of the models was tested by applying homography transformations with varying degrees of severity (max_offset = 1.5, 2.0, 2.5). As shown in Figure 4, the accuracy of all models decreases as

the severity of the transformation increases. CNN outperformed other models, demonstrating its ability to handle perspective distortions more effectively.

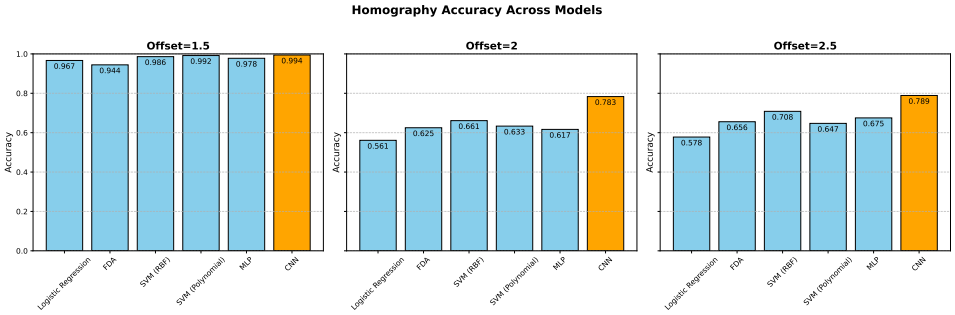


Figure 4: Accuracy under homography transformations with different max_offset values.

5.3 Impact of Gaussian Noise

Gaussian noise with varying sigma values (0.2, 0.5, 0.7) was added to the test data to simulate noisy environments. The results, presented in Figure 5, reveal that the models' performance degrades as the noise level increases. However, SVM with a polynomial kernel and CNN demonstrated relatively high tolerance to Gaussian noise.

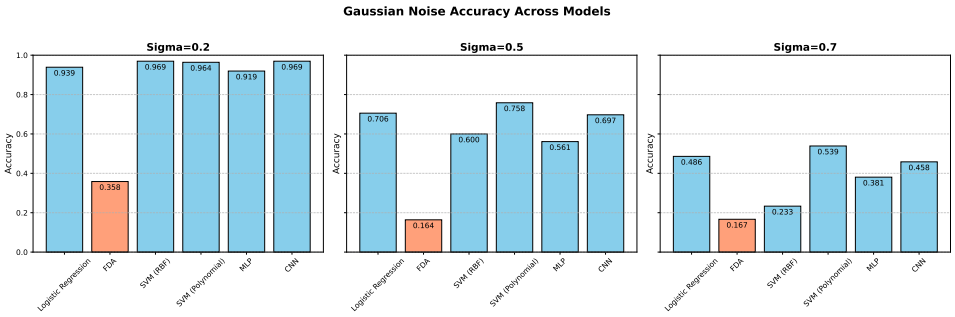


Figure 5: Accuracy under Gaussian noise transformations with different sigma values.

5.4 Analysis and Comparison of Results

5.4.1 Brief Explanation of Algorithms

Each algorithm has unique characteristics that make it suitable for specific tasks:

- **Logistic Regression**: A simple linear classifier effective for baseline evaluations.
- **Fisher Discriminant Analysis (FDA)**: Projects data to maximize class separability but struggles with non-linear boundaries.
- **Support Vector Machine (SVM)**: Excels in finding decision boundaries with kernels for linear and non-linear tasks.
- **Multilayer Perceptron (MLP)**: A basic neural network capable of learning non-linear patterns.
- **Convolutional Neural Network (CNN)**: Best-suited for image-based tasks due to its ability to learn spatial hierarchies.

5.4.2 Performance Analysis for Each Algorithm

The results reveal that CNN consistently outperforms other models, achieving the highest accuracy under both baseline and transformation scenarios. However, the following patterns were observed:

- **Baseline Accuracy**: CNN and SVM (Polynomial Kernel) delivered the best results, demonstrating their capability to model complex patterns in data.
- **Homography Transformation**: CNN showed higher resilience to perspective shifts, maintaining relatively high accuracy even under severe offsets.
- **Gaussian Noise**: SVM with Polynomial Kernel and CNN were the most robust to noise, maintaining stable accuracy compared to other models.

5.4.3 Comparative Results

Table 2 provides a summary of the comparative results across all scenarios, showcasing the models' performance under varying conditions. This comparison highlights the trade-offs between linear and non-linear models and demonstrates the superiority of CNN in handling real-world distortions.

Table 2: Comparative Results for Each Algorithm Across All Scenarios

Model	Baseline Accuracy	Homography (Avg.)	Gaussian Noise (Avg.)
Logistic Regression	0.9667	0.7018	0.7102
FDA	0.9444	0.7417	0.2293
SVM (RBF Kernel)	0.9861	0.7851	0.6009
SVM (Polynomial Kernel)	0.9917	0.7574	0.7537
MLP	0.9778	0.7565	0.6203
CNN	0.9944	0.8555	0.7083

5.4.4 Pros and Cons of Each Algorithm

Finally, the strengths and weaknesses of each model are summarized below:

- **Logistic Regression**:
  - **Pros**: Simple and efficient for linear tasks.
  - **Cons**: Limited by its inability to handle non-linear boundaries.
- **Fisher Discriminant Analysis (FDA)**:
  - **Pros**: Maximizes class separability for linearly separable data.
  - **Cons**: Poor performance on non-linear and noisy data.
- **SVM (RBF Kernel)**:
  - **Pros**: Excels at modeling non-linear decision boundaries.
  - **Cons**: Requires careful parameter tuning and is computationally intensive.
- **SVM (Polynomial Kernel)**:
  - **Pros**: Performs well on both linear and non-linear data.
  - **Cons**: Susceptible to overfitting with complex kernels.
- **Multilayer Perceptron (MLP)**:
  - **Pros**: Capable of learning non-linear relationships.
  - **Cons**: Requires substantial training data and is sensitive to hyperparameters.
- **Convolutional Neural Network (CNN)**:
  - **Pros**: Best suited for image data, handles transformations well.
  - **Cons**: Computationally expensive and requires a large amount of labeled data.

6 Conclusion

In this study, we evaluated the robustness of six machine learning models (Logistic Regression, FDA, SVM with RBF and Polynomial kernels, MLP, and CNN) under two common types of data transformations: homography and Gaussian noise. These transformations simulate real-world distortions, such as changes in perspective and sensor noise, commonly encountered in autonomous driving perception systems.

Key Findings

- **Baseline Performance**: CNN and SVM with Polynomial Kernel achieved the highest accuracy on the original MNIST dataset, demonstrating their capacity to model complex patterns effectively.
- **Homography Transformation**: CNN exhibited superior resilience to perspective changes, maintaining high accuracy even under the most severe transformation ($max\_offset = 2.5$). SVM with RBF kernel also performed relatively well in this scenario.
- **Gaussian Noise**: Models like CNN and SVM with Polynomial Kernel demonstrated strong tolerance to noise, whereas FDA struggled significantly, highlighting its limitation in handling noisy environments.

Practical Implications

The findings of this study provide valuable insights into selecting appropriate models for image-based tasks in noisy or distorted environments:

- For tasks involving significant image distortions (e.g., perspective shifts), CNN is the most reliable choice.
- For tasks with moderate noise levels, both CNN and SVM with Polynomial Kernel are recommended.
- Simpler models like Logistic Regression and FDA may be suitable for clean, well-prepared data but are less robust under challenging conditions.

Limitations and Future Work

While this study provides a comprehensive evaluation of model robustness, several limitations remain:

- The study was conducted on a simplified dataset (MNIST). Future work should explore more complex datasets that better represent real-world scenarios.
- Additional transformations, such as scaling, rotation, and motion blur, could be evaluated to expand the scope of the analysis.
- Advanced techniques like data augmentation and adversarial training could be investigated to improve model robustness further.

In conclusion, this study highlights the importance of selecting robust machine learning models for real-world applications, particularly in autonomous driving systems, where data distortions are unavoidable. By understanding the strengths and limitations of various models, practitioners can make informed decisions to ensure reliable performance under diverse conditions.