Comparative Analysis of Feature Extraction Methods for Signature Verification

Ibrahim Akhtar
Department of Data Science

ibrahimakhtar3@gmail.com

***Abstract*—This paper presents a comprehensive comparison of different feature extraction techniques for signature verification. I implement and evaluate three approaches: a Convolutional Neural Network (CNN) for end-to-end feature learning, Histogram of Oriented Gradients (HOG), and Scale-Invariant Feature Transform (SIFT) combined with Support Vector Machine (SVM) classifiers. The methods are evaluated on a signature verification dataset containing genuine and forged signatures from multiple subjects. Performance metrics including accuracy, precision, recall, and F1-score are used to assess each approach. Our results demonstrate that while the CNN model achieves competitive performance with an accuracy of 77.58%, the HOG-SVM combination outperforms both other methods with 87.88% accuracy. The findings highlight the effectiveness of traditional feature engineering techniques when combined with appropriate classifiers for signature verification tasks, particularly when training data is limited.**

***Index Terms*—signature verification, convolutional neural networks, histogram of oriented gradients, scale-invariant feature transform, support vector machines, biometrics**

I. Introduction

Signature verification is a fundamental biometric authentication technique with applications in banking, legal documents, and secure access systems. The task involves determining whether a given signature belongs to a claimed individual by comparing it with reference signatures. A reliable signature verification system must be able to handle variations in genuine signatures while detecting forgeries.

Traditional approaches to signature verification rely on manual feature extraction techniques that capture the geometric and stylistic characteristics of signatures. More recently, deep learning methods have shown promising results by automatically learning discriminative features directly from raw signature images. However, it remains unclear whether these deep learning approaches consistently outperform traditional feature engineering methods across different datasets and conditions.

In this paper, I compare three distinct approaches to signature verification:

- A Convolutional Neural Network (CNN) [2] that learns features directly from the raw signature images
- Histogram of Oriented Gradients (HOG) [3] features with Support Vector Machine (SVM) [4] classifiers
- Scale-Invariant Feature Transform (SIFT) [5] descriptors with SVM classifiers

I evaluate these methods on a publicly available signature verification dataset, focusing on their ability to correctly identify the signer from their signature. Our analysis provides insights into the strengths and limitations of each approach and offers recommendations for practical implementations of signature verification systems.

II. Methodology

A. Dataset

For this research, I used the Signature Verification Dataset from Kaggle [1], which contains genuine and forged signatures from multiple individuals. The dataset is organized with each signer having a separate folder containing their genuine signatures, and a corresponding folder with forged signatures (denoted by the suffix ”_forg”).

B. Data Preprocessing

All signature images were preprocessed using the following steps:

- Conversion to grayscale to reduce dimensionality and computational complexity
- Resizing to a standard dimension of 150×150 pixels to ensure uniform input size
- Normalization of pixel values to the range [0, 1] to improve training stability

The dataset was split into training (80%) and validation (20%) sets, with stratification to ensure balanced representation of each signer's signatures in both sets.

C. Model Architectures

1) CNN Architecture: I designed a CNN architecture with the following structure:

- Input normalization layer to standardize images
- Three convolutional blocks, each containing:
  - Two convolutional layers with 3×3 kernels
  - Max pooling layer with 2×2 pool size
  - Dropout (0.25) for regularization
- Flatten layer to convert the 2D feature maps to a 1D feature vector
- Dense layer with 1024 neurons and ReLU activation
- Dropout (0.5) for regularization
- Output layer with softmax activation, with one neuron per signature class

The model was trained using the AdamW optimizer with a learning rate of 0.0001 and weight decay of 1e-5. I used categorical cross-entropy as the loss function and trained for up to 20 epochs with early stopping based on validation loss.

2) HOG Feature Extraction: For the HOG-based approach, I extracted HOG features with the following parameters:

- Window size: 150 × 150 pixels (same as image size)
- Cell size: 8 × 8 pixels
- Block size: 16 × 16 pixels (2 × 2 cells)
- Block stride: 8 × 8 pixels (50% overlap between blocks)
- Number of orientation bins: 9

These HOG features were then used to train an SVM classifier with a radial basis function (RBF) kernel.

3) SIFT Feature Extraction: For the SIFT-based approach, I:

- Detected SIFT keypoints and computed descriptors for each image
- Computed the mean of all descriptors for each image to create a fixed-length feature vector
- Used these feature vectors to train an SVM classifier with an RBF kernel

D. Evaluation Metrics

To evaluate and compare the performance of each approach, I used the following metrics:

- Accuracy: The proportion of correctly classified signatures
- Precision: The proportion of true positive predictions among all positive predictions
- Recall: The proportion of true positive predictions among all actual positives
- F1-score: The harmonic mean of precision and recall

III. Results

A. Performance Comparison

Table I shows the performance metrics for each of the three approaches on the validation set.

TABLE I: Performance Comparison of Different Feature Extraction Methods

Model	Accuracy	Precision	Recall	F1-Score
CNN	0.7758	0.8057	0.7758	0.7725
HOG-SVM	0.8788	0.9052	0.8788	0.8743
SIFT-SVM	0.2485	0.1762	0.2485	0.1626

As shown in Table I, the HOG-SVM approach achieved the highest performance across all metrics, with an accuracy of 87.88% and an F1-score of 87.43%. The CNN approach achieved competitive results with 77.58% accuracy and 77.25% F1-score. The SIFT-SVM approach performed significantly worse, with only 24.85% accuracy.

Fig. 1 shows a graphical comparison of the performance metrics for the three approaches. The HOG-SVM combination clearly outperforms the other methods across all metrics, while SIFT-SVM shows the poorest performance.

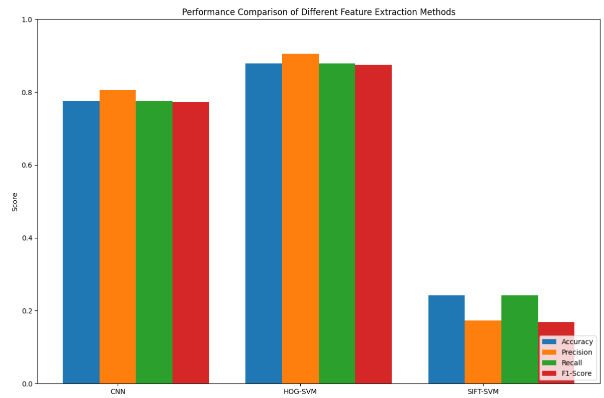

Fig. 1: Performance comparison of different feature extraction methods.

B. Confusion Matrices

Figs. 2, 3, and 4 show the confusion matrices for the CNN, HOG-SVM, and SIFT-SVM approaches, respectively. These matrices provide a detailed view of the classification performance across different signature classes.

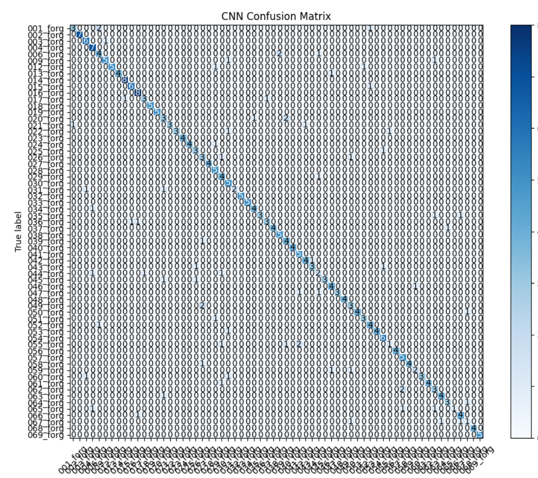

Fig. 2: Confusion matrix for the CNN model.

IV. Discussion

A. Comparison of Feature Extraction Methods

The results clearly demonstrate that the HOG-SVM approach outperforms both the CNN and SIFT-SVM approaches for this signature verification task. This finding is somewhat surprising given the recent trend toward deep learning methods for computer vision tasks. There are several possible explanations for this result:

- **Dataset size**: The dataset may not be large enough for the CNN to learn robust features effectively. CNNs typically require large amounts of training data to generalize well.
- **Feature representation**: HOG features capture local gradient information that is particularly relevant for signature

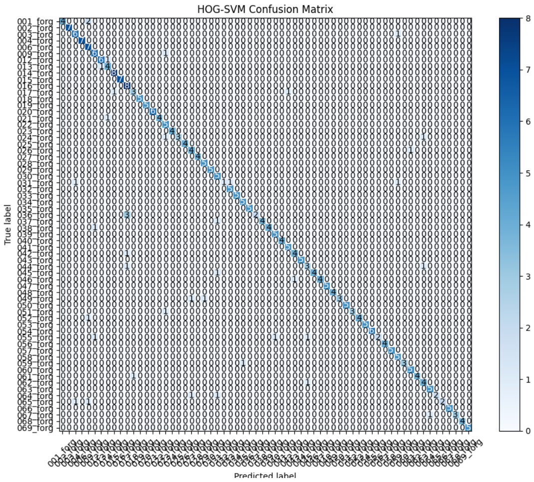

Fig. 3: Confusion matrix for the HOG-SVM model.

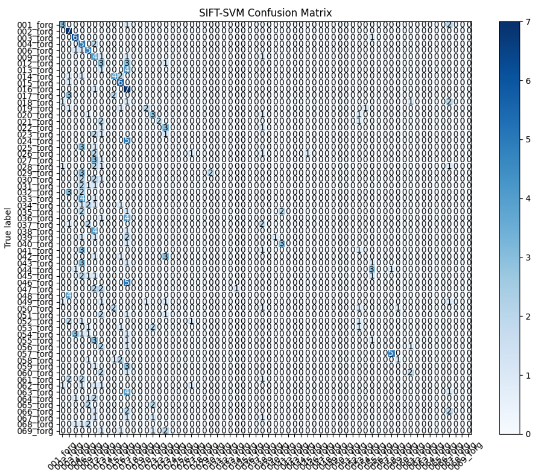

Fig. 4: Confusion matrix for the SIFT-SVM model.

verification, where stroke directions and patterns are key discriminative features.

- **Model complexity**: The CNN model, with its millions of parameters, may be prone to overfitting on this dataset, despite the use of dropout and early stopping.

The poor performance of SIFT-SVM is likely due to the variability in the number of keypoints detected in different signature images and the simplistic approach of taking the mean of all descriptors, which may lose important spatial information.

B. Analysis of Confusion Matrices

The confusion matrices reveal some interesting patterns:

- The HOG-SVM confusion matrix shows a strong diagonal pattern, indicating good classification performance across most classes.
- The CNN confusion matrix also shows a clear diagonal pattern but with more off-diagonal elements, suggesting more misclassifications compared to HOG-SVM.
- The SIFT-SVM confusion matrix shows a much weaker diagonal pattern, with classifications distributed more randomly across classes, confirming its poor performance.

These patterns suggest that both HOG-SVM and CNN are learning meaningful signature features, with HOG-SVM capturing more discriminative features for this specific task.

C. Practical Implications

The superior performance of HOG-SVM suggests that traditional feature engineering approaches still have significant value in specialized computer vision tasks like signature verification. While deep learning methods offer end-to-end feature learning, they may not always be the optimal choice, especially when:

- Training data is limited
- The task benefits from specific, hand-crafted features
- Computational efficiency is important

For practical signature verification systems, a hybrid approach combining hand-crafted features with deep learning might provide the best performance.

V. Conclusion

In this paper, I compared three different approaches to signature verification: CNN, HOG-SVM, and SIFT-SVM. Our findings indicate that HOG features combined with SVM classifiers outperform both the CNN and SIFT-SVM approaches on our dataset, achieving 87.88% accuracy compared to 77.58% for CNN and 24.85% for SIFT-SVM.

These results highlight the continued relevance of traditional feature engineering techniques in computer vision tasks, particularly when domain knowledge can guide the selection of appropriate features. However, the competitive performance of the CNN suggests that with larger datasets or more sophisticated architectures, deep learning approaches could potentially match or exceed the performance of traditional methods.

Future work could explore ensemble methods combining multiple feature extraction techniques, more sophisticated deep learning architectures, or the use of one-shot or few-shot learning approaches to handle limited training data more effectively.

References

[1] R. Reni, "Signature Verification Dataset," Kaggle, 2020. [Online]. Available: https://www.kaggle.com/datasets/robinreni/signature-verification-dataset

[2] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," Proceedings of the IEEE, vol. 86, no. 11, pp. 2278–2324, 1998.

[3] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), 2005, pp. 886–893.

[4] C. Cortes and V. Vapnik, "Support-vector networks," Machine Learning, vol. 20, no. 3, pp. 273–297, 1995.

[5] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," International Journal of Computer Vision, vol. 60, no. 2, pp. 91–110, 2004.

[6] M. Abadi et al., "TensorFlow: Large-scale machine learning on heterogeneous systems," 2015. [Online]. Available: http://tensorflow.org/

[7] G. Bradski, "The OpenCV Library," Dr. Dobb's Journal of Software Tools, 2000.