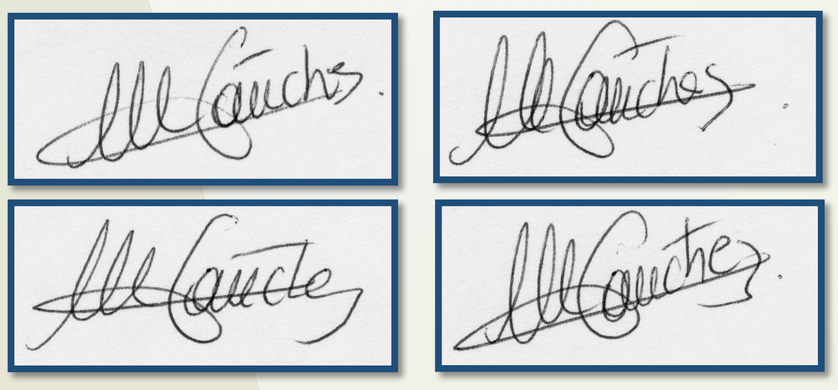
OFFLINE SIGNATURE VERIFICATION USING MACHINE LEARNING

Burak Mert Gonultas
Advisor : Assoc. Prof. Hazim Kemal Ekenel

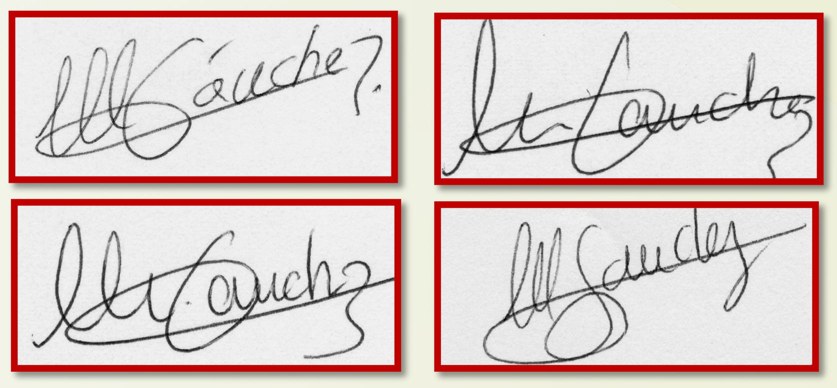
Istanbul Technical University
gonultasbu@itu.edu.tr
github.com/gonultasbu/sigver_v2

INTRODUCTION

Genuine Signatures



Forged Signatures



- Feature representations generated by CNNs which were trained only using genuine signatures of different users have been explored.
- Feature representation vectors have been classified using Support Vector Machines, which were trained only using genuine signatures.
- Classification performances were measured using genuine and skilled forgery signatures, similar to examples presented above.
- Both syntethic and biometric databases have been utilized for training and performance measurement.
- Considerable performance increase has been achieved compared to previous work in the academic literature

METHODOLOGY

1. Feature representations
   - Generated using a trained CNN very similar to AlexNet
   - Vectors of 2048 features
2. SVM Classification
   - Binary classifiers learning user signature representations
   - Nonlinear RBF kernel
   - Learning from low number of genuine samples

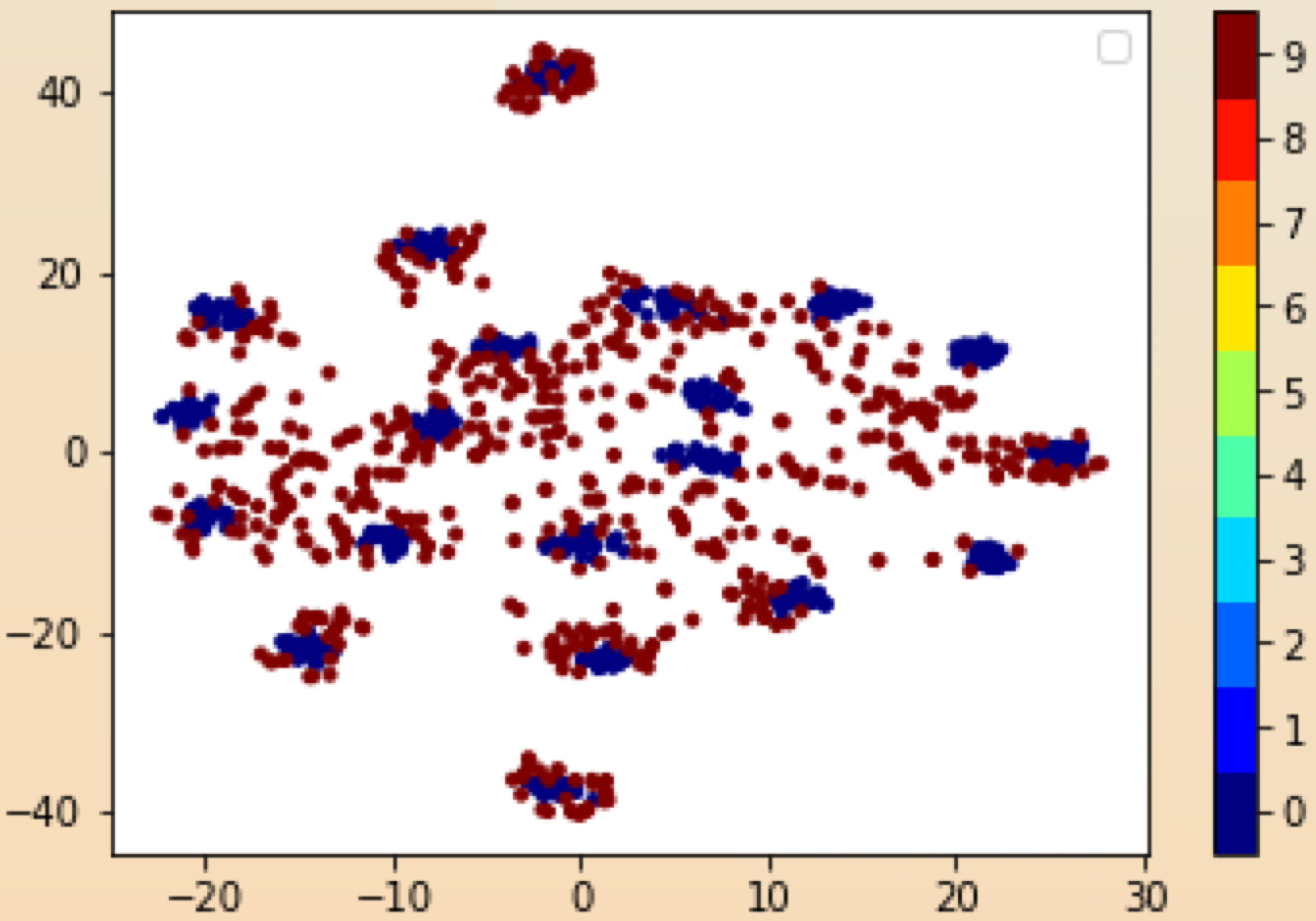


Figure 1: Scatter plot generated using t-SNE representation of 20 randomly sampled users' feature representations, red dots represent forgeries while the blue dots represent genuine signatures. Other colors in legend can be ignored for this representation.

Table 1: CNN Architecture

Layer	Size	Parameters
Input	1x150x220	
Convolution	96x11x11	stride = 4
Max. Pool	96x3x3	stride = 2
Convolution	256x5x5	stride = 1, pad = 2
Max. Pool	256x3x3	stride = 2
Convolution	384x3x3	stride = 1, pad = 1
Convolution	384x3x3	stride = 1, pad =1
Convolution	256x3x3	stride = 1, pad =1
Max. Pool	256x3x3	stride = 2
Fully Connected	2048	
Fully Connected	2048	

EVALUATION

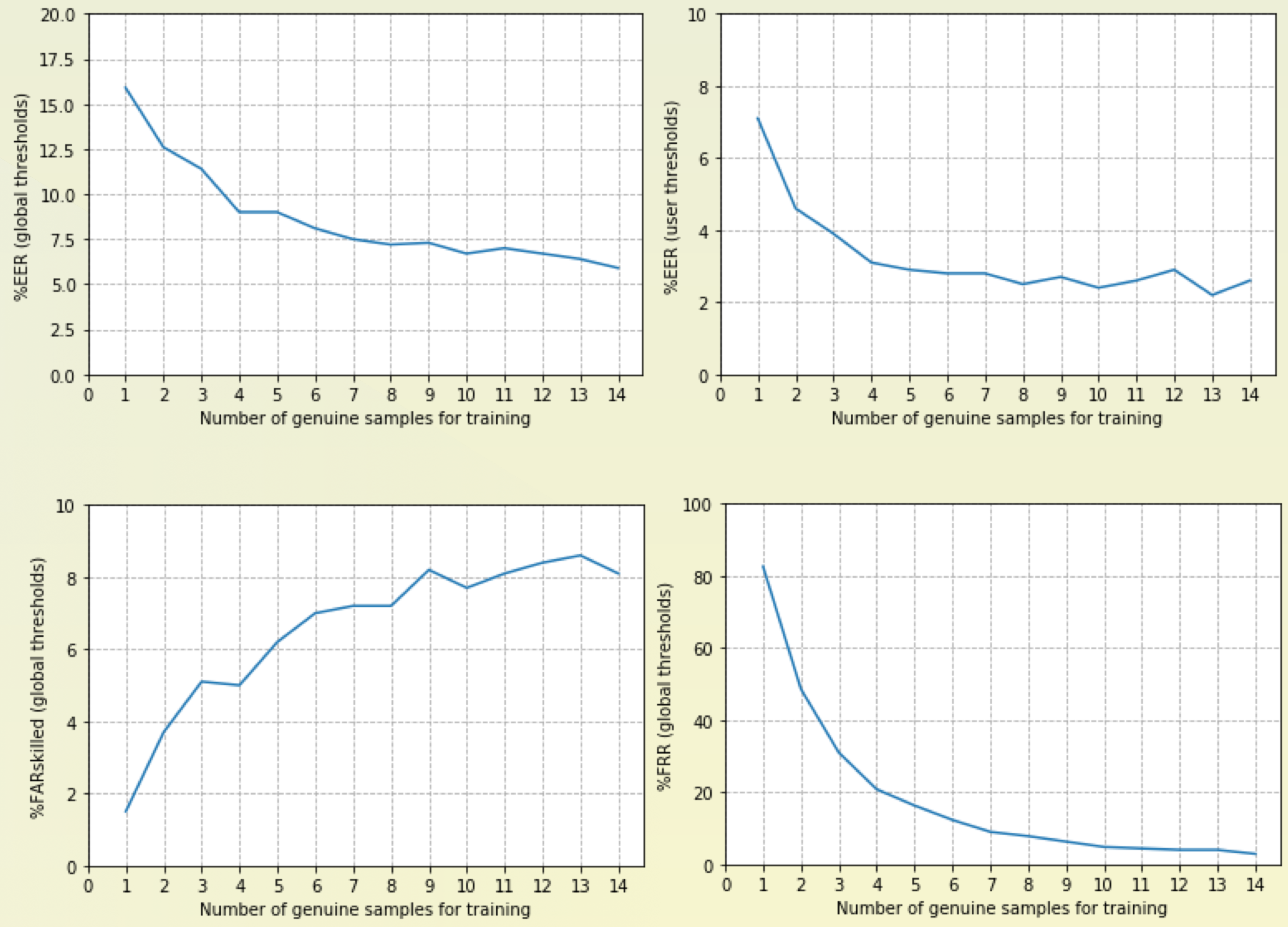


Figure 2: Sensitivity to genuine sample count used during training.

Table 2: Performance on GPDS-Synthetic 4000, using 10 genuine, 10 skilled forgery signatures.

Reference	EER_{global}(%)
Dutta et al.	26.33
Dey et al.	22.24
Soleimani et al.	13.30
Ferrer et al.	16.44
Serdouk et al.	16.68
Zhang et al.	14.79
Proposed	**6.96**

Table 3: Performance on MCYT-75, using 10 genuine, 10 skilled forgery samples.

Reference	EER_{user} (%)
Gilperez et al.	6.44
Vargas et al.	7.08
Ooi et al.	9.87
Soleimani et al.	9.86
Hafemann et al.	**2.87**
Proposed	4.00

Table 4: Performance on GPDS using pre-extracted features from Hafemann et al.

Reference	# genuine samples	EER_{user} (%)
Hu and Chen	10	7.66
Guerbai et al.	12	15.07
Serdouk et al.	16	12.52
Soleimani et al.	10	20.94
Yilmaz.	12	6.97
Hafemann et al.	10	1.69
Proposed	**10**	**1.33**

CONCLUSION

- A comprehensive study on offline signature verification using handwritten signatures using CNN and SVM based approaches has been presented.
- The impact of genuine sample count has been investigated using a fixed decision threshold, which also contained information about robustness against varying genuine sample counts.
- Generalization capability of features learned on synthetic dataset has been measured.
- The learnable visual cues that make a signature a skilled forgery are proven to be too dataset specific.