2018 CVPR
SALT LAKE CITY • JUNE 18-22

Hybrid User-Independent and User-Dependent Offline Signature Verification with a Two-Channel CNN

Mustafa Berkay Yılmaz (Assist. Prof., Akdeniz University)
berkayyilmaz@akdeniz.edu.tr

Kağan Öztürk (MS student, Akdeniz University)
kaganozturk1992@gmail.com

- **Introduction**
- Preprocessing
- Two channel CNN
- User-independent (UI) verification
- User-dependent (UD) verification
- Concurrent UI / UD verification
- Related works
- Experimental results

Introduction

Static / offline / handwritten signature verification task

Random forgery

Simple forgery

Skilled forgery

- Introduction
- **Preprocessing**
- Two channel CNN
- User-independent (UI) verification
- User-dependent (UD) verification
- Concurrent UI / UD verification
- Related works
- Experimental results

Preprocessing

Invert the gray-level values

Eliminate small connected components

Preprocessing

Align query to reference during training (brute-force)

Try scale, rotation and translation combinations to the query

Use basic LBP feature as alignment match metric

Reference

Not so slow: All transformations are applied to references ahead of time, only compared with the query.
Inverse of best transform match is applied to the query

Resize to 100 x 150 and input to the CNN

- Introduction
- Preprocessing
- **Two channel CNN**
- User-independent (UI) verification
- User-dependent (UD) verification
- Concurrent UI / UD verification
- Related works
- Experimental results

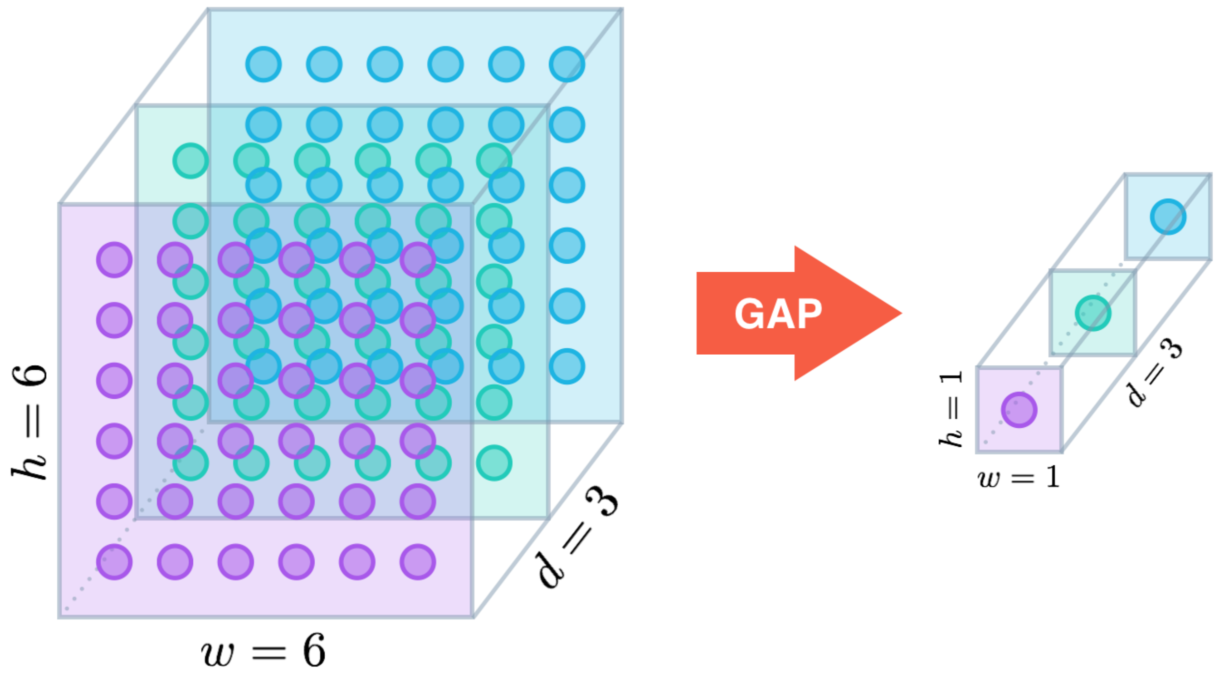
Two-channel CNN

Learn how to decide if a query signature Q is genuine or not

In existence of a reference R (known to be genuine)

With the help of a two-channel CNN $\phi(R, Q)$

- replace max-pooling layers by convolutional layers of increasing stride
- 5 dropout layers of probability 0.5
- Use Global Average Pooling before fully-connected layer

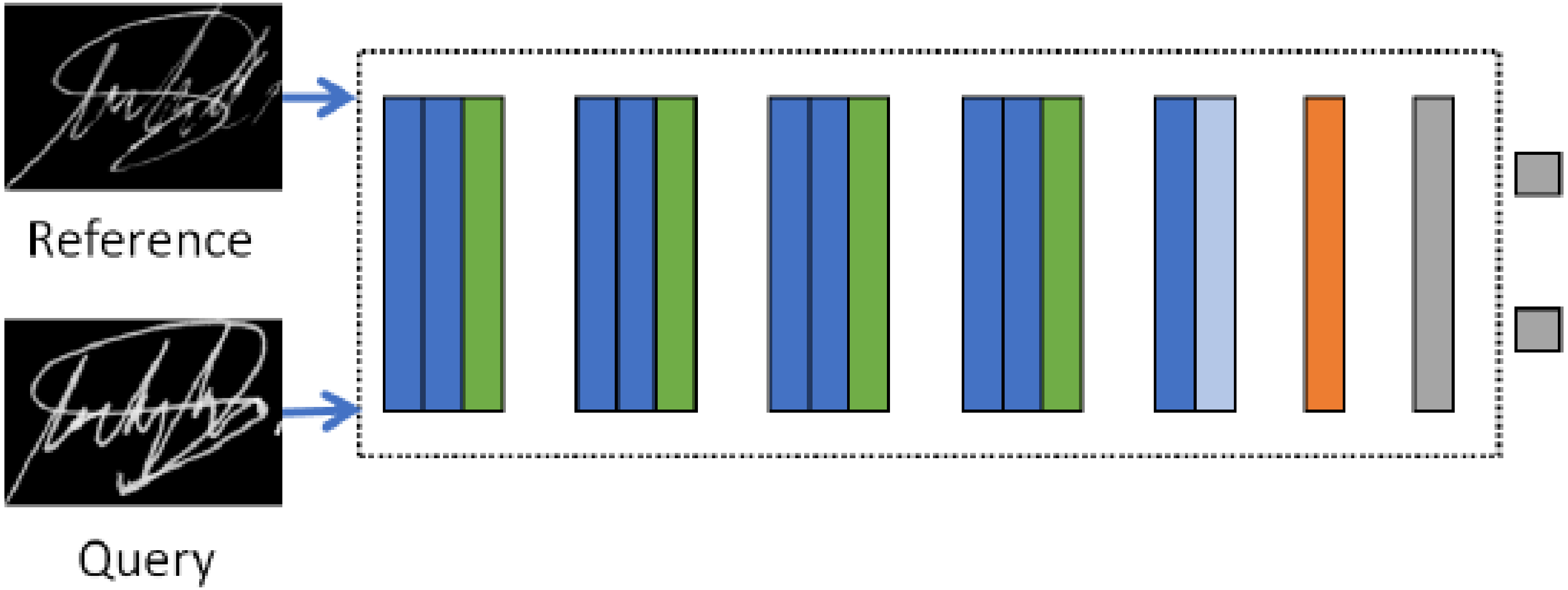


https://alexisbcook.github.io/2017/global-average-pooling-layers-for-object-localization/

Two-channel CNN

- Conv layer filter size 3 stride 1
- Conv layer filter size 3 stride 2 in gray-level model, max pooling in binary model
- Conv layer filter size 1 stride 1
- Global average pooling layer
- Fully connected layer

Input size is
100 x 150 x 2



- Introduction
- Preprocessing
- Two channel CNN
- **User-independent (UI) verification**
- User-dependent (UD) verification
- Concurrent UI / UD verification
- Related works
- Experimental results

User-independent (UI) verification

Probability $P(y|R_n^y, Q)$ of a query signature Q belonging to user y in the presence of a reference signature $R_n^y \in \mathbf{R}^\mathbf{y}$

(reference set of user y)

Estimated using CNN: $P(y|R_n^y, Q) \approx \phi(R_n^y, Q)$

Calculate the average score over references: $P_{ui}(y|\mathbf{R}^\mathbf{y}, Q) \approx \sum_{n=1}^{N} \phi(R_n^y, Q)/N$

- No user-specific model has to be trained and stored
- No concern of model update when a user provides new reference signatures
- When the number of reference signatures is 1, can still obtain effectual verification score

- Introduction
- Preprocessing
- Two channel CNN
- User-independent (UI) verification
- **User-dependent (UD) verification**
- Concurrent UI / UD verification
- Related works
- Experimental results

User-dependent (UD) verification

Training UD classifiers

Use signature representations obtained as the output of the GAP layer before the fully-connected layer

(with a reference and a query as the input)

We have as many representations for a query Q as the number of references

Feature set becomes $F_Q = \cup_{n=1}^{N} \phi_{GAP}(R_n^y, Q)$

(Dimensionality is 200 after the GAP layer)

User-dependent (UD) verification

- Utilize SVM with RBF kernel to train UD models
- All N x (N - 1) genuine-genuine inter-reference pairs as positive samples (2nd reference pretends a genuine query)
- Genuine-forgery pairs from **other subjects** are randomly selected as negative samples
  - We can assume that we have some training subjects for whom we have both genuine and forgery samples

User-dependent (UD) verification

During testing, we have N different representations for an unknown query signature Q, so we have N SVM scores. Take the average SVM score:

$$P_{ud}(y|\mathbf{R}^y, Q) \approx \sum_{n=1}^{N} f^y(\phi_{GAP}(R_n^y, Q))/N$$

SVM decision function of user y

- Introduction
- Preprocessing
- Two channel CNN
- User-independent (UI) verification
- User-dependent (UD) verification
- **Concurrent UI / UD verification**
- Related works
- Experimental results

Concurrent UI/UD verification

Score level fusion of UI and UD classifiers corresponds to a classifier combination of UI neural net and UD SVM

$$P_{uid}(y|\mathbf{R}^{\mathbf{y}},Q) = \alpha P_{ui}(y|\mathbf{R}^{\mathbf{y}},Q) + (1-\alpha)P_{ud}(y|\mathbf{R}^{\mathbf{y}},Q)$$

Learn the weight from a validation set

- Introduction
- Preprocessing
- Two channel CNN
- User-independent (UI) verification
- User-dependent (UD) verification
- Concurrent UI / UD verification
- **Related works**
- Experimental results

Related works

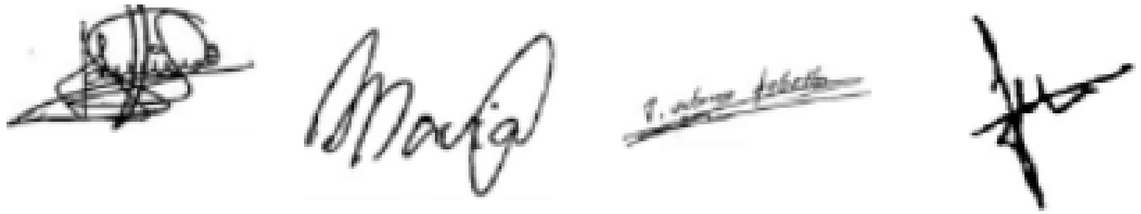
Experimental protocol variations in different works:

- Databases
- Train & test subsets
- Image formats (gray-level or binary)
- Selection & number of reference samples
- Using skilled forgeries in training or not
- Using random forgeries in testing or not
- Hyper-parameter selection
- Calculation of decision thresholds

Databases

GPDS960 signature database

[1] M. Blumenstein, Miguel A. Ferrer, J.F. Vargas, “The 4NSigComp2010 off-line signature verification competition: Scenario 2”, in proceedings of 12th International Conference on Frontiers in Handwriting Recognition, ISSBN: 978-0-7695-4221-8, pp. 721-726, Kolkata, India, 16-18 November 2010.



MCYT baseline corpus

[2] Javier Ortega-Garcia, J Fierrez-Aguilar, D Simon, J Gonzalez, M Faundez-Zanuy, V Espinosa, A Satue, I Hernaez, J-J Igarza, C Vivaracho, D Escudero, Q-I Moro, "MCYT baseline corpus: a bimodal biometric database," in *IEE Proceedings - Vision, Image and Signal Processing*, vol. 150, no. 6, pp. 395-401, 15 Dec. 2003.

Related works

[3] EER of 7%, with **binary** GPDS-160 subset using 12 reference signatures per subject. Combination of handcrafted feature classifiers

[4] EER of 7.21%, **gray-level** GPDS-300 using 5 references. (2.70% with ideal user-based thresholds). Sparse dictionary learning and coding

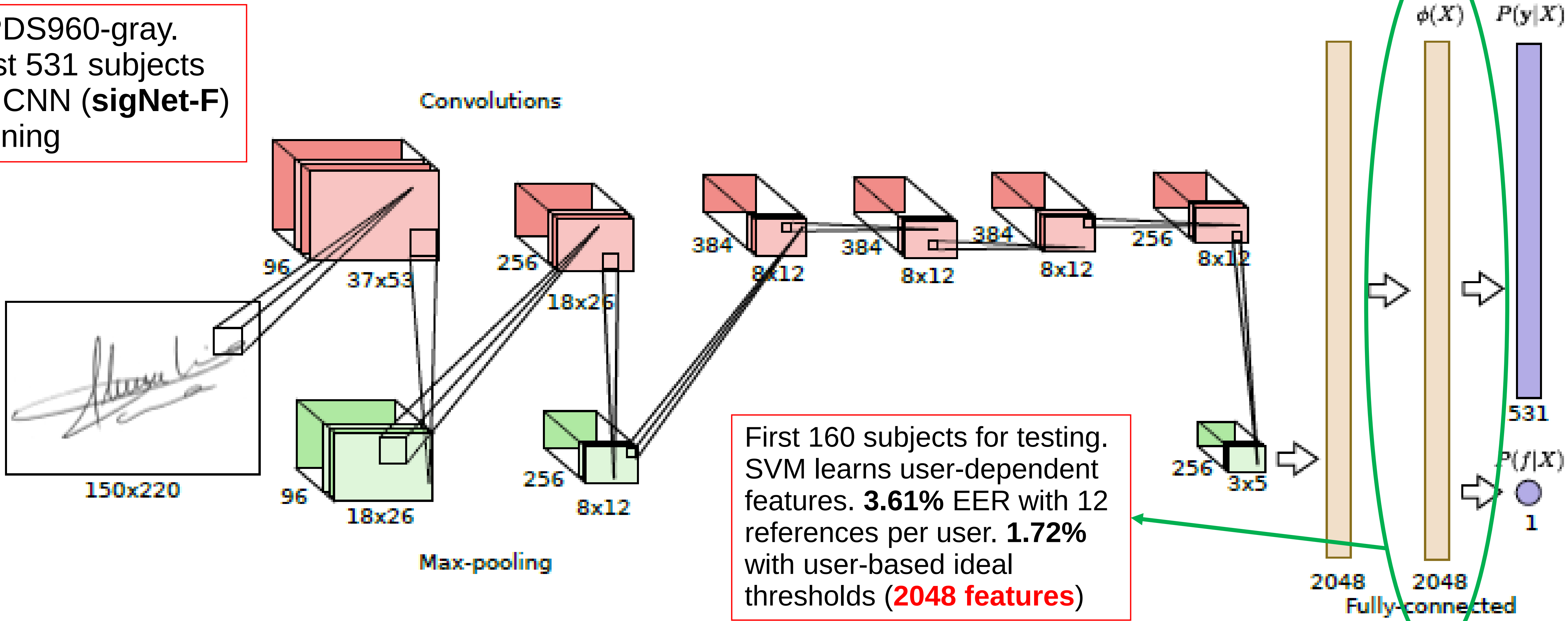
[5] EER of 20% with 5 reference signatures using **binary** GPDS-160. Signature representation is learnt by PCANet (a basic deep learning structure) from a separate set of users

[3] M. B. Yılmaz and B. Yanıkoğlu. Score level fusion of classifiers in off-line signature verification. Information Fusion, 32(Part B):109 – 119, 2016. SI Information Fusion in Biometrics.
[4] E. N. Zois, I. Theodorakopoulos, D. Tsourounis, and G. Economou. Parsimonious coding and verification of offline handwritten signatures. In 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), pages 636–645, July 2017.
[5] M. B. Yılmaz. Offline signature verification with user-based and global classifiers of local features. PhD thesis, Sabancı University, 2015.

Related works

[6] User-independent signature image representation learning using CNN.

GPDS960-gray.
Last 531 subjects for CNN (**sigNet-F**) training



First 160 subjects for testing. SVM learns user-dependent features. **3.61%** EER with 12 references per user. **1.72%** with user-based ideal thresholds (**2048 features**)

[6] L. G. Hafemann, R. Sabourin, and L. S. Oliveira. Learning features for offline handwritten signature verification using deep convolutional neural networks. Pattern Recogn., 70(C):163–176, Oct. 2017.

- Introduction
- Preprocessing
- Two channel CNN
- User-independent (UI) verification
- User-dependent (UD) verification
- Concurrent UI / UD verification
- Related works
- **Experimental results**

Experimental results

- Error measures
  - Equal error rate (EER): When false accept (FA) and false reject (FR) rates are equal
  - Distinguishing error rate (DER): Average of FA and FR
- Database
  - GPDS960-gray (881 users, 24 genuine samples and *at most* 30 forgery samples)
  - We investigate the sensitivity of the proposed method to gray-level and binary signature images
  - Manually converted into binary

Experimental protocol

- IDs inclusive [460 - 960] for CNN training **(t)**
- [358 - 459] for CNN validation **(V1)**
- [205 - 357] for UD SVM grid search **(V2)**
- **V2** for selecting the UI+UD combination weight
- **V2** for selecting the combination weights of our final UI+UD score and UD scores obtained in [6]

Experimental protocol

- [2 - 204] for test set **T** (reference (genuine) samples **T1** for UD training, rest of the samples **T2** for UI and UD testing)
  no skilled forgeries in UD training
  - Consider N=1, 5 and 12; **2** equal partitions of genuine samples: 12 for test.
  - Randomly select N samples **3** times for actual reference set: **6** random repetitions for each N
- Genuine-forgery signature pair representations from *other* users of T as negatives

Experimental protocol

- Calculation of EER in 3 different ways:
  - Directly from test set (global threshold)
  - Directly from test set, using normalized subject scores (user-based thresholds)
  - Learning the threshold from V2

    (FR and FA may now be different, we use DER in this case)

Experimental protocol

Separation of the database into subsets

T_1 T_2	V_2	V_1	τ
160 subjects	146 subjects	100 subjects	475 subjects

Samples

Subjects

Results

Results with **gray-level** t and V1 for **UI and UD** (V2 threshold results excluded)

	N	Global threshold(T) EER		User-based thresholds EER	
		UI	UD	UI	UD
Gray V_2 and T	1	$8.74 \pm 0.34\%$	-	$6.81 \pm 0.17\%$	-
	5	$7.39 \pm 0.22\%$	$6.52 \pm 0.68\%$	$5.75 \pm 0.75\%$	$4.72 \pm 0.33\%$
	12	$7.20 \pm 0.24\%$	$4.29 \pm 0.14\%$	$5.78 \pm 0.67\%$	$2.88 \pm 0.18\%$
Binary V_2 and T	1	$32.74 \pm 0.44\%$	-	$29.74 \pm 0.64\%$	-
	5	$31.92 \pm 0.31\%$	$23.49 \pm 0.65\%$	$27.26 \pm 0.35\%$	$19.65 \pm 0.42\%$
	12	$31.22 \pm 0.42\%$	$17.95 \pm 0.50\%$	$26.80 \pm 1.07\%$	$15.03 \pm 0.21\%$

Results

Results with **gray-level** t and V1 for the combination of UI and UD

	N	Global threshold (V_2) DER	Global threshold(T) EER	User-based thresholds EER
Gray V_2 and T	5	$5.23 \pm 0.21\%$	$5.38 \pm 0.14\%$	$3.92 \pm 0.28\%$
	12	$4.82 \pm 0.06\%$	$4.13 \pm 0.31\%$	$2.94 \pm 0.28\%$
Binary V_2 and T	5	$40.68 \pm 0.45\%$	$21.57 \pm 0.35\%$	$18.21 \pm 0.46\%$
	12	$20.81 \pm 0.75\%$	$18.08 \pm 0.43\%$	$14.73 \pm 0.02\%$

Results with **binary** t and V1 for the combination of UI and UD

	N	Global threshold (V_2) DER	Global threshold(T) EER	User-based thresholds EER
Gray V_2 and T	5	$14.20 \pm 0.43\%$	$14.10 \pm 0.32\%$	$10.85 \pm 0.39\%$
	12	$13.86 \pm 0.23\%$	$11.12 \pm 0.29\%$	$8.26 \pm 0.08\%$
Binary V_2 and T	5	$23.30 \pm 0.55\%$	$15.40 \pm 0.35\%$	$11.31 \pm 0.21\%$
	12	$12.15 \pm 0.13\%$	$11.86 \pm 0.02\%$	$9.22 \pm 0.15\%$

Results

UD results with the features extracted using SigNet-F CNN [6]

N	Global threshold (V_2) DER	Global threshold (T) EER	User-based thresholds EER
5	$5.81 \pm 0.63\%$	$4.44 \pm 0.19\%$	$2.66 \pm 0.40\%$
12	$3.82 \pm 0.55\%$	$3.66 \pm 0.58\%$	$2.08 \pm 0.64\%$

(Dimensionality of 200)

Score-level combination results of (gray-gray) **two-channel CNN final score** with **SigNet-F UD** (Dimensionality of 2048)

N	Global threshold (V_2) DER	Global threshold (T) EER	User-based thresholds EER
5	$2.90 \pm 0.31\%$	$2.33 \pm 0.17\%$	$1.16 \pm 0.21\%$
12	$1.75 \pm 0.36\%$	$1.76 \pm 0.37\%$	$0.88 \pm 0.36\%$

- Thank you
- Questions