

Identity Verification from Human Scent using Learned Representations of Two-Dimensional Gas Chromatography–Mass Spectrometry Data

Jan Hlavsa ^{1*}, Radim Spetlik ¹, Jana Čechová ², Petra Pojmanová ²,
Jiří Matas ¹, Štěpán Urban ²

¹Faculty of Electrical Engineering, Czech Technical University in Prague, Prague, Czech Republic.

²Department of Analytical Chemistry, University of Chemistry and Technology, Prague, Prague, Czech Republic.

*Corresponding author(s). E-mail(s): hlavsja3@fel.cvut.cz;

Supplementary Information

Machine Vision and Applications

S1 Additional Methodological Details and Experimental Results

This supplementary material extends the supplementary information accompanying our WACV 2026 conference paper [8] and retains most of its methodological descriptions and benchmark results. The description of Channel Representation Learning, previously presented in the conference paper, is included here for completeness, whereas the section on Alternative Correlation-Based Representations is new to the journal extension.

We first describe the GC×GC–ToF-MS instrumentation, the compound peak detectors introduced in [3], and the canonical-frame registration procedure. We then provide details

on learning the channel representation, including the alternative triangular kernel formulation, and describe the architecture of our custom CNN. Further sections present comprehensive results for the baseline method developed by analytical chemists [1], deterministic descriptors, the comparison between the two mass spectrometers, regularization and data augmentation, and the complete evaluation of the CNN embedding encoders. Finally, we investigate several alternative correlation-based representations, including temporal and m/z Gram matrices, spatial correlation tensors, and learned projections for raw-pixel aggregation.

S1.1 Sensor Instrumentation

Analyses were conducted on an Agilent 7890B gas chromatograph (Agilent, Santa Clara, CA, USA) coupled to time-of-flight mass spectrometers Pegasus[®] 4D-C and Pegasus[®] BT-4D (LECO Corp., St. Joseph, MI, USA). Automated sample introduction was performed with a MultiPurpose Sampler MPS Robotic (Gerstel, Mülheim, Germany). Raw data were exported from the ChromaTOF[®] software (versions

4.72.0.0, 5.51.06.0, and 5.55.41.BT; LECO Corp., St. Joseph, MI, USA).

The GC $\times$ GC-ToF-MS configuration followed a previously reported layout [6] employing a reverse column arrangement: a medium-polar primary column Rtx-200MS (Restek, Bellefonte, PA, USA, 30 m $\times$ 0.25 mm $\times$ 0.25 μ m) and a non-polar secondary column TG-5HT (Thermo Fisher Scientific, Waltham, MA, USA, 1 m $\times$ 0.25 mm $\times$ 0.25 μ m). Of the total secondary-column length, 10 cm were located in the modulator, 69 cm in the secondary oven, and 21 cm in the GC-MS transfer line. A 2 m pre-column of the same stationary phase as the primary column was installed. When using the Pegasus BT-4D detector, the secondary column totaled 1.1 m to accommodate 31 cm within the GC-MS transfer line.

Prior to injection, samples were homogenized by shaking for 10 min followed by 10 min in an ultrasonic bath. Injections of 1 μ L were made with a 1 μ L air plug in the syringe, at 280°C in splitless mode (2 min splitless time). Carrier gas was operated at a constant-flow rate of 1.5 mL/min. The oven program was: 40°C (2 min), ramp at 5°C/min to 320°C (10 min), for a total runtime of 68 min (4080 s). The secondary oven was maintained at +5°C relative to the primary oven and the modulator at +20°C relative to the primary. Cryogenic modulation of the secondary column used dried air at -80°C. The run comprised three modulation segments: 500-1704 s (period 6 s; hot pulse 1.8 s; cold pulse 1.2 s), 1704-2592 s (period 8 s; hot 2.4 s; cold 1.6 s), and 2592-4080 s (period 10 s; hot 3.3 s; cold 1.7 s).

The transfer line was held at 280°C and the ion source at 250°C. Electron ionization (EI) at 70 eV was used. Total ion current (TIC) data were collected over m/z 29–800 at 200 spectra s⁻¹. The acquisition delay was 500 s, and the detector voltage was set 200 V above the tuned value to enhance sensitivity.

S1.2 Compound Peak Detectors

We locate the characteristic peak of each compound using an annotated 4D-C dataset of 40 individuals (504 total samples) from [3, 5]. Although we train our compound-peak detectors on a small annotated set, we subsequently rectify the entire Human Scent Dataset (Sec. 4 of the main paper)

using these learned detectors, eliminating the need for extensive annotation by chemical experts. To align the spectral resolutions, we quantized the high-spectral-resolution measurements to match the dimensions of the low-spectral-resolution data. We then stored the detected peak positions and applied the registration procedure to each channel representation using these stored positions.

S1.2.1 Chemical Compounds Being Detected

In the registration step, we followed [3] exactly. We used the set of 22 chemical compounds from the paper and also experimented with the addition of two more compounds, Heptacosane and Nonacosane. See the list of compounds in Table S1. Note that a standardized mass spectrum of all listed compounds is available in public databases.

#	Compound
1	Tetradecane, 1-chloro-
2	2,5-pyrrolidinedione, 3-decyl-
3	2,5-pyrrolidinedione, 3-dodecyl-
4	2,5-pyrrolidinedione, 3-tetradecyl-
5	2,5-pyrrolidinedione, 3-hexadecyl-
6	Squalene
7	Cholesterol
8	Cholesta-3,5-dien-7-one
9	Cholest-4-en-3-one
10	Cholesta-4,6-dien-3-one
11	Hexadecanoic acid, dodecyl ester
12	9-Hexadecenoic acid, tetradecyl ester
13	Hexadecanoic acid, tetradecyl ester
14	9-Hexadecenoic acid, hexadecyl ester
15	Hexadecanoic acid, hexadecyl ester
16	9-Hexadecenoic acid, octadecyl ester
17	Hexadecanoic acid, octadecyl ester
18	1-Dodecanol
19	Dodecane, 1-chloro-
20	Dodecanamide, N-allyl-
21	Fumaric acid, 2-dimethylaminoethyl heptyl ester
22	Cholesta-3,5-diene (isomer 2)
23	Heptacosane
24	Nonacosane

Table S1 List of compounds used for registration in our experiments, following [3]. The first 22 compounds correspond to the set actually used for registration in [3]. Heptacosane and Nonacosane were only mentioned in the original paper but not used there; in our experiments, we evaluated both the 22-compound and 24-compound variants. Reproduced from the conference version [8]

S1.2.2 Bounding Box Constraint

Many compounds share similar mass spectra (Fig. 1 of the main paper), so *positional cues* are crucial as discussed in [3]. We restrict the search for compound c to a bounding box around its average position $\bar{\mathbf{u}}_c$:

$$\text{BBox}(\bar{\mathbf{u}}_c, w, h) = \left\{ \mathbf{u} \in \Omega : \begin{aligned} |\bar{u}_{1,c} - u_1| &< \frac{w}{2}, \\ |\bar{u}_{2,c} - u_2| &< \frac{h}{2} \end{aligned} \right\}. \quad (\text{S1})$$

where $\Omega := \{1, \dots, H\} \times \{1, \dots, W\}$ is the pixel grid, and w, h are the bounding-box dimensions.

S1.2.3 Weighted Cosine Peak Detector

Cosine similarity alone can yield spurious matches. To reduce those, we apply a spatial penalty that was presented in [3]:

$$\hat{\mathbf{u}}_c^{\text{ccos}}(\bar{\mathbf{u}}_c, \bar{s}_c) = \arg \max_{\mathbf{u} \in \text{BBox}(\bar{\mathbf{u}}_c)} \left[\cos(\bar{s}_c, s_{\mathbf{u}}) - p(\bar{\mathbf{u}}_c, \mathbf{u}) \right].$$

where

$$p(\bar{\mathbf{u}}_c, \mathbf{u}) = (|\bar{u}_{1,c} - u_1| + 6|\bar{u}_{2,c} - u_2|)^4.$$

The exponent $(\cdot)^4$ strongly penalizes large deviations, and the factor of 6 accounts for typically wider shifts along the first chromatographic dimension. This detector thus prefers spectra near $\bar{\mathbf{u}}_c$ that match $\bar{s}_c$.

Cosine similarity measures the angular alignment between two spectra, making the approach invariant to absolute intensity scaling.

S1.2.4 Fully Convolutional Network (FCN) Peak Detector

A fully convolutional network (FCN) that was shown in [3] predicts *class likelihood* at each pixel. Let Ω_c be the set of pixels classified by the FCN as belonging to compound c . We combine the FCN prediction with cosine similarity:

$$\hat{\mathbf{u}}_c^{\text{FCN}(\mathbf{u}, s)_{\text{cos}}^+} = \arg \max_{\mathbf{u} \in \Omega_c} \left(\text{FCN}(\mathbf{u}, s) + \cos(\bar{s}_c, s_{\mathbf{u}}) \right).$$

In both detectors, the final peak location is the pixel $\mathbf{u}$ that maximizes the respective score.

S1.3 Canonical Frame Registration

As noted in [3], “variations in instrument conditions (e.g., temperature changes, aging, injection timing) shift compound retention times and lead to misalignment of GC×GC–ToF–MS measurements.” Canonical-frame registration was therefore considered a key step in their processing pipeline. We reimplemented the procedure introduced in [3] and conducted experiments with and without registration to assess its impact in our setting.

Let $\mathcal{R}^{\text{ref}} = \{c_1, c_2, \dots\}$ denote the set of reference compounds. For each $c \in \mathcal{R}^{\text{ref}}$, let $\mathbf{u}_c \in \mathbb{R}^2$ be the reference location and s_c^{ref} the corresponding mass spectrum.

In each human scent sample (GC×GC–ToF–MS measurement) $\mathcal{S}$, we first detect the occurrences (see Sec. S1.2 for detectors description) of a fixed set of reference compounds (see Sec. S1.2.1 for the list). This set corresponds to the compounds that were reported as consistent across samples in [3]. For each compound $c \in \mathcal{R}^{\text{ref}}$ we obtain its detected position $\mathbf{u}_c^{(\mathcal{S})} \in \mathbb{R}^2$ and corresponding mass spectrum $s_c^{(\mathcal{S})}$.

Let

$$\mathcal{U}^{\text{ref}} = \{\mathbf{u}_c \mid c \in \mathcal{R}^{\text{ref}}\}$$

denote the reference locations. Given the detected positions

$$\mathcal{U}^{(\mathcal{S})} = \{\mathbf{u}_c^{(\mathcal{S})} \mid c \in \mathcal{R}^{\text{ref}}\},$$

we construct a Delaunay triangulation of $\mathcal{U}^{(\mathcal{S})}$. Each resulting triangle is then mapped to the triangle formed by the corresponding points in $\mathcal{U}^{\text{ref}}$. For each triangle in the Delaunay triangulation, we compute a local affine mapping that transforms the detected vertices $\mathbf{u}_{c_1}^{(\mathcal{S})}, \mathbf{u}_{c_2}^{(\mathcal{S})}, \mathbf{u}_{c_3}^{(\mathcal{S})}$ of sample $\mathcal{S}$ to the corresponding reference vertices $\mathbf{u}_{c_1}, \mathbf{u}_{c_2}, \mathbf{u}_{c_3}$. A point $\mathbf{u}^{(\mathcal{S})}$ in the measured sample is first expressed in barycentric coordinates $(\lambda_1, \lambda_2, \lambda_3)$ with respect to $(\mathbf{u}_{c_1}^{(\mathcal{S})}, \mathbf{u}_{c_2}^{(\mathcal{S})}, \mathbf{u}_{c_3}^{(\mathcal{S})})$,

$$\begin{aligned} \mathbf{u}^{(\mathcal{S})} &= \lambda_1 \mathbf{u}_{c_1}^{(\mathcal{S})} + \lambda_2 \mathbf{u}_{c_2}^{(\mathcal{S})} + \lambda_3 \mathbf{u}_{c_3}^{(\mathcal{S})}, \\ \lambda_1 + \lambda_2 + \lambda_3 &= 1, \\ \lambda_i &\geq 0. \end{aligned}$$

The canonical position $\mathbf{u}$ is then obtained by mapping the same barycentric coordinates to the reference triangle,

$$\mathbf{u} = \lambda_1 \mathbf{u}_{c_1} + \lambda_2 \mathbf{u}_{c_2} + \lambda_3 \mathbf{u}_{c_3}.$$

Finally, the entire human scent sample is interpolated on the transformed mesh using piecewise linear interpolation, yielding the canonical frame representation.

This process allows comparison of scent samples taken under different conditions as was introduced in [3].

S1.4 Channel Representation Learning

To obtain a compact channel representation of the high-resolution mass spectra, we learn the parameters of the kernel bank by reconstructing the original spectral intensities. The following describes the reconstruction model and the corresponding training objective.

Model.

For a pixel $\mathbf{u}$ with spectrum $\{(m_{n,\mathbf{u}}, I_{n,\mathbf{u}})\}_{n=1}^{N(\mathbf{u})}$, we approximate intensities with kernel basis κ and $\mathbf{C}_{k,\mathbf{u}}$

$$\hat{I}_{n,\mathbf{u}} = \sum_{k=1}^C \mathbf{C}_{k,\mathbf{u}} \kappa(m_{n,\mathbf{u}}; \theta_k). \quad (\text{S2})$$

Objective.

Parameters $\{\alpha_k, \theta_k\}_{k=1}^C$ are learned by minimizing the weighted MSE over training pixels Ω :

$$E = \sum_{\mathbf{u} \in \Omega} \sum_{n=1}^{N(\mathbf{u})} \frac{1}{w_{n,\mathbf{u}}} (I_{n,\mathbf{u}} - \hat{I}_{n,\mathbf{u}})^2. \quad (\text{S3})$$

We use either uniform $w_{n,\mathbf{u}} = 1$ or inverse-intensity

$$w_{n,\mathbf{u}} = \max(I_{\Omega}/100, I_{n,\mathbf{u}}), \quad I_{\Omega} = \max_{\mathbf{u} \in \Omega, n} I_{n,\mathbf{u}}. \quad (\text{S4})$$

After training, $\{\alpha_k, \theta_k\}$ are frozen and reused for all pixels.

S1.5 Triangular Kernel Function For Channel Representation

In addition to the Gaussian kernel function introduced in Sec. 3.2 of the main paper, we experimented with triangular kernel

$$\kappa_{\Delta}(m_n; \mu_k, w_k) = \max(1 - \frac{|m_n - \mu_k|}{w_k}, 0),$$

a piecewise linear function whose half-width $w_k > 0$ is learned. In this case, the kernel centers μ_k are set using the same heuristic as in Sec. 3.2 of the main paper.

We apply a 10-fold cross-validation on the dataset of 2528 samples by 497×2000 retention times by “number of spectral samples”. For the Triangular kernel, we again minimize the weighted reconstruction error E (S3) using stochastic gradient descent and compare two weight schemes: uniform weighting and inversed-intensity weighting (S3).

κ	w	$C = 3$	$C = 5$	$C = 10$	$C = 20$
G	U	$6.3 \cdot 10^{-8}$	$1.4 \cdot 10^{-7}$	$4.5 \cdot 10^{-7}$	$3.9 \cdot 10^{-8}$
G	I	$5.1 \cdot 10^{-6}$	$1.7 \cdot 10^{-5}$	$5.1 \cdot 10^{-7}$	$8.0 \cdot 10^{-7}$
Δ	U	$2.7 \cdot 10^{-5}$	$1.7 \cdot 10^{-4}$	$2.5 \cdot 10^{-5}$	$2.1 \cdot 10^{-5}$
Δ	I	$1.5 \cdot 10^{-4}$	$1.5 \cdot 10^{-4}$	$1.6 \cdot 10^{-4}$	$1.0 \cdot 10^{-6}$

Table S2 Mean squared reconstruction error of channel representation with C kernels over 10-fold cross-validation of Human Scent Dataset (Sec. 4 of the main paper) with two kernel types κ (Gaussian kernel G and Triangular kernel Δ) and two weight types (uniform U and inversed intensity I). Reporting only μ . Reproduced from the conference version [8]

The results are presented in Tab. S2. The reconstruction error of triangular kernels is $\approx$ two orders of magnitude higher than the error of Gaussian kernels. We attribute this to the higher expressiveness and smoother approximation capability of the Gaussian kernel. Therefore, all results in the main paper are reported using Gaussian kernels.

S1.6 Architecture of Custom CNN

As a custom-designed feature extractor, we introduce a compact convolutional network, to realize the embedding function f . The model comprises three modules: (i) a convolutional backbone, (ii)

	Metric	Threshold τ_{feat}		
		not set	$\tau_{\text{feat}} = 0.1$	$\tau_{\text{feat}} = 0.5$
AUC	Euclidean	62.1 ± 6.5	49.6 ± 3.3	49.2 ± 2.6
	Pearson	57.7 ± 7.4	56.2 ± 6.7	55.2 ± 4.2
	Spearman	79.3 ± 6.2	56.1 ± 5.2	54.8 ± 2.9
TPR@5%FPR	Euclidean	13.7 ± 9.3	5.7 ± 2.0	3.7 ± 1.6
	Pearson	10.8 ± 9.7	9.8 ± 4.9	6.6 ± 2.1
	Spearman	42.0 ± 12.6	5.8 ± 3.5	1.4 ± 2.0

Table S3 Cross-validation Area Under ROC Curve and TPR@5%FPR of baseline method developed by analytical chemists [1] on 4D-C dataset [3] Reproduced from the conference version [8]

a down-sampling stack, and (iii) a residual MLP head.

Backbone consists of three stages, each featuring: (i) a 2D Convolution block with a kernel 2×8 , stride (1, 2), and padding (0, 1), (ii) a Residual 2D Convolution Block that enhances feature propagation, and (iii) dropout layers with probabilities scaled by the dropout strength (i.e., 0.1, 0.2 and 0.3).

Down-sampling Module applies four sequential 2D Convolution block layers using a 5×5 kernel, stride 2, and padding 1, each followed by dropout at $0.3 \times$ dropout strength to progressively reduce spatial dimensions.

Residual MLP Head maps the high-dimensional feature vector, the flattened down-sampled features, to the desired space representation, with an additional dropout of $0.5 \times$ dropout strength to mitigate overfitting.

(Residual) 2D Convolution Block sequentially applies a 2D convolution, 2D batch normalization, a ReLU activation, and dropout. In its residual variant, the block adds the input element-wise to its output. When the input and output channel dimensions differ, a depth-wise 2D convolution is optionally applied to the input to ensure proper dimensionality alignment.

S1.7 Evaluation of Baseline Method Developed by Analytical Chemists

The full results of the experiment from Sec. 3.3.1 of the main paper are presented in Table S3.

S1.8 Evaluation of Deterministic Descriptors

In this section, we evaluate three deterministic feature descriptors, LBP, HOG, and SIFT, and systematically analyze how their hyperparameter configurations influence verification performance. Detailed results are reported in Table S4.

For LBP, we vary the neighbourhood radius, $r \in \{1, 2, 3\}$, and examine both the *basic* (default) and *uniform* mapping schemes. The resulting feature embedding vector is a normalized histogram over the LBP codes.

- **Basic / default mapping.** The full P -bit code obtained by thresholding the P neighbours against the centre pixel is retained unchanged, producing a histogram with 2^P bins [4]. This variant preserves maximum detail but yields a high-dimensional, potentially sparse representation.
- **Uniform mapping.** Codes that contain more than two circular bit transitions $0 \leftrightarrow 1$ are merged into a single “miscellaneous” class, whereas the 58 uniform patterns (for $P=8$) remain distinct. The resulting 59-bin histogram reduces dimensionality by a factor of approximately seven and is empirically more robust to noise and small rotations.

For HOG, the feature embedding is a flattened vector formed by concatenating contrast-normalized histograms of gradient orientations computed across the image. We systematically vary three hyperparameters:

- **Cell size.** Each cell is a square patch of either 8×8 or 16×16 pixels in the input image.
- **Number of orientation bins.** The histogram can contain 6, 9, or 12 equally spaced bins.
- **Block geometry c .** For contrast normalization, neighbouring cells are grouped into rectangular blocks of either 2×2 or 3×3 cells.

For SIFT, we vary four key hyperparameters while maintaining the default number of layers per octave (3). The final feature vector is constructed by computing the mean of all local 128-dimensional SIFT descriptors, resulting in a fixed-length global representation. Specifically, we explore the following hyperparameters:

	parameters				$c_{\cos}(\bar{\mathbf{u}}_c, \bar{\mathbf{s}}_c)$				$\text{FCN}(\mathbf{u}, \mathbf{s})_{\cos}^+$					
					no registration		$\mathcal{R}^{(22)}$		$\mathcal{R}^{(24)}$		$\mathcal{R}^{(22)}$		$\mathcal{R}^{(24)}$	
	θ_1	θ_2	θ_3	θ_4	AUC	TPR@5%	AUC	TPR@5%	AUC	TPR@5%	AUC	TPR@5%	AUC	TPR@5%
HOG	8x8	6	3x3		72.4±7.3	39.6±8.1	74.3±7.2	39.1±7.8	74.4±7.3	39.3±7.7	75.8±4.3	33.0±7.2	76.6±4.5	33.1±7.1
	8x8	6	2x2		73.0±6.6	39.1±7.8	73.4±6.5	38.9±7.6	73.5±6.6	38.9±7.7	75.3±3.6	31.9±6.8	76.2±3.7	31.8±7.1
	16x16	6	3x3		71.0±8.1	40.5±8.4	74.2±7.8	39.8±8.1	74.3±7.9	39.9±8.2	75.2±4.5	33.0±7.1	75.8±4.8	32.8±7.1
	8x8	9	3x3		72.2±7.9	39.4±8.0	73.9±7.9	38.8±8.0	74.0±7.9	39.3±8.1	76.0±4.9	33.4±7.2	76.7±5.0	33.5±7.1
	16x16	6	2x2		71.4±7.6	39.1±7.8	73.5±7.5	39.1±7.8	73.6±7.5	39.2±7.8	75.3±4.3	32.0±6.8	76.2±4.5	31.7±6.8
	8x8	9	2x2		72.9±7.1	38.5±7.5	73.3±7.1	38.2±7.5	73.3±7.2	38.5±7.8	75.7±4.0	32.2±6.8	76.5±4.2	32.2±6.9
	16x16	9	3x3		71.2±8.5	40.1±8.3	73.9±8.2	39.5±8.1	74.0±8.3	39.6±8.2	75.2±4.9	33.1±7.0	75.8±5.2	33.2±7.2
	8x8	12	2x2		72.6±7.4	38.0±7.5	73.1±7.4	37.7±7.7	73.2±7.5	38.1±7.8	75.7±4.3	32.5±6.8	76.6±4.5	32.7±6.8
	8x8	12	3x3		72.0±8.2	39.1±7.9	73.8±8.1	38.6±8.0	73.9±8.2	39.0±8.2	76.0±5.1	33.8±7.3	76.7±5.3	33.8±7.1
	16x16	9	2x2		71.8±7.8	38.8±7.7	73.5±7.8	38.9±7.9	73.5±7.9	39.1±8.0	75.6±4.5	32.0±6.8	76.4±4.8	32.1±6.7
	16x16	12	3x3		71.3±8.6	39.9±8.3	73.8±8.3	39.3±8.0	73.9±8.4	39.5±8.2	75.2±5.1	33.5±7.1	75.8±5.3	33.5±7.1
	16x16	12	2x2		71.8±8.0	38.6±7.7	73.5±7.9	38.9±7.9	73.5±8.0	39.1±8.0	75.6±4.7	32.5±7.0	76.4±4.9	32.3±6.7
LBP	1	d			78.1±3.2	15.8±2.2	77.1±2.8	15.8±2.2	77.1±2.8	15.5±2.1	75.6±3.0	16.2±2.0	75.9±3.0	16.0±2.4
	1	u			76.9±2.9	17.6±2.3	75.7±3.1	18.9±3.2	75.7±3.2	19.0±3.3	76.8±4.0	22.4±4.3	78.1±4.3	22.9±4.9
	2	u			70.0±3.3	19.8±2.2	72.8±3.4	20.8±3.1	72.9±3.4	20.7±3.2	73.1±5.0	22.2±3.7	73.7±5.5	21.9±4.1
	3	u			67.8±3.4	20.1±2.9	70.9±2.9	20.9±3.0	71.0±3.0	21.1±3.1	71.3±5.2	21.3±3.6	71.1±5.8	20.9±3.9
	3	d			71.2±4.9	16.4±3.1	70.0±5.1	16.8±3.2	70.0±5.1	16.7±3.2	64.3±4.0	15.9±2.0	64.3±3.8	15.9±1.9
	2	d			67.6±2.8	13.2±1.4	65.6±2.9	12.3±1.1	65.7±2.9	12.4±1.0	61.3±3.3	11.6±1.2	61.4±3.4	10.9±1.6
SIFT	1000	0.04	20	1.6	64.1±4.6	15.3±3.6	63.8±4.1	15.5±4.0	63.4±3.9	15.1±3.8	59.0±6.7	13.4±4.0	58.2±6.4	13.1±4.0
	0	0.04	20	1.6	64.1±4.6	15.3±3.6	63.8±4.1	15.5±4.0	63.4±3.9	15.1±3.8	59.0±6.7	13.4±4.0	58.2±6.4	13.1±4.0
	1000	0.04	20	1.2	61.2±5.6	14.7±3.9	61.1±5.0	14.9±4.2	60.8±5.1	14.6±4.0	56.7±6.6	12.3±4.1	56.6±6.2	12.5±3.8
	0	0.04	20	1.2	61.2±5.6	14.7±3.9	61.1±5.0	14.9±4.2	60.8±5.1	14.6±4.0	56.7±6.6	12.3±4.1	56.6±6.2	12.5±3.8
	1000	0.04	10	1.6	53.2±5.8	12.3±3.8	59.2±5.3	14.4±3.2	59.1±5.2	14.5±3.3	56.1±7.5	13.9±4.2	55.5±7.1	13.3±4.1
	0	0.04	10	1.6	53.2±5.8	12.3±3.8	59.2±5.3	14.4±3.2	59.1±5.2	14.5±3.3	56.1±7.5	13.9±4.2	55.5±7.1	13.3±4.1
	1000	0.01	20	1.6	60.1±7.3	17.2±5.3	59.1±7.3	16.8±5.5	58.9±7.4	16.8±5.2	58.8±7.0	15.5±4.6	59.4±6.9	15.6±4.7
	0	0.01	20	1.6	60.1±7.3	17.1±5.3	59.2±7.3	16.6±5.3	59.0±7.4	16.6±5.1	59.2±7.0	15.5±4.6	60.1±6.9	16.1±5.0
	1000	0.01	20	1.2	63.7±7.2	17.9±5.2	60.1±7.3	16.8±4.8	60.1±7.3	16.7±4.8	61.4±6.7	15.4±4.1	62.1±6.4	15.8±3.8
	0	0.01	20	1.2	63.8±7.1	17.8±5.2	60.7±7.2	16.8±4.7	60.6±7.2	16.7±4.7	62.7±6.7	16.5±4.9	63.8±6.5	17.4±4.9
	1000	0.01	10	1.2	66.6±6.0	18.6±5.3	58.7±7.0	16.8±5.2	58.5±6.9	16.3±5.1	58.4±7.2	15.1±4.8	58.5±7.2	15.5±4.8
	0	0.01	10	1.2	66.6±6.0	18.5±5.3	58.7±7.0	16.8±5.1	58.5±6.9	16.3±5.0	58.5±7.2	15.0±4.7	58.8±7.2	16.0±5.0
	1000	0.04	10	1.2	55.0±5.8	13.2±3.4	57.4±5.5	13.6±3.6	57.2±5.7	13.4±3.4	54.6±7.2	12.5±3.7	54.4±6.8	12.6±4.1
	0	0.04	10	1.2	55.0±5.8	13.2±3.4	57.4±5.5	13.6±3.6	57.2±5.7	13.4±3.4	54.6±7.2	12.5±3.7	54.4±6.8	12.6±4.1
	0	0.01	10	1.6	56.7±8.1	17.1±5.5	56.1±7.3	16.7±5.2	56.0±7.4	16.5±5.1	55.6±7.3	15.4±4.8	56.1±7.4	16.2±5.5
	1000	0.01	10	1.6	56.7±8.1	17.1±5.5	56.1±7.3	16.7±5.2	56.0±7.4	16.5±5.1	55.6±7.3	15.4±4.8	56.1±7.4	16.2±5.5
NN					85.8±5.1	39.0±11.8	88.0±3.1	48.3±9.2	88.0±3.6	50.5±10.5	87.9±3.5	49.1±8.9	87.4±2.9	50.2±9.1

Table S4 Validation AUC and TPR@5%FPR for non-registered and registered data and LBP, HOG, and SIFT descriptors, and our custom CNN using single-channel data and 10-fold cross-validation. Results are reported as mean $\pm$ standard deviation across folds. Best AUC in bold, best TPR@5%FPR in italic bold. Parameters: LBP: θ_1 = radius, θ_2 = method (u = uniform, d = default); HOG: θ_1 = pixels per cell, θ_2 = number of orientation bins, θ_3 = cells per block; SIFT: θ_1 = number of features, θ_2 = contrast threshold, θ_3 = edge threshold, θ_4 = sigma. Reproduced from the conference version [8]

- **Number of features** $n \in \{0, 1000\}$. This parameter specifies the maximum number of keypoints retained after detection; $n=0$ disables the cap, whereas $n=1000$ limits the descriptor set to the 1000 strongest responses.
- **Contrast threshold** $f \in \{0.001, 0.04\}$. Filters out weak features in semi-uniform regions.
- **Edge threshold** $e \in \{10, 20\}$. Filters out edge-like features.

S1.9 Comparison of High- and Low-resolution Mass Spectrometers

We evaluate our custom CNN separately on the high- and low-resolution subsets of HSD (Sec. 4 of the main paper). For each detector, identity- and gender-aware folds are used, containing only samples from that detector. The cohorts are *not identity-paired* (only 3 identities overlap). High-resolution results are reported as $\mu \pm \sigma$ over 10 folds; low-resolution rows report $\Delta = \mu_{\text{low}} - \mu_{\text{high}}$ (with σ omitted). See Table S5.

			$c_{\cos}(\bar{\mathbf{u}}_c, \bar{\mathbf{s}}_c)$						$\text{FCN}(\mathbf{u}, \mathbf{s})_{\cos}^+$					
			no registration		$\mathcal{R}^{(22)}$		$\mathcal{R}^{(24)}$		$\mathcal{R}^{(22)}$		$\mathcal{R}^{(24)}$			
κ	w	C	AUROC	TPR	AUROC	TPR	AUROC	TPR	AUROC	TPR	AUROC	TPR		
High-resolution time-of-flight mass spectrometer Pegasus® BT 4D														
custom	CNN	G	I	3	89.2±3.6	57.2±8.0	90.8±3.4	57.1±9.6	91.0±2.8	58.7±8.4	80.6±5.1	37.8±9.4	80.9±5.3	35.8±9.0
custom	CNN	G	I	5	88.5±3.5	54.6±6.3	86.6±4.3	51.9±6.3	87.4±4.1	52.0±6.8	77.7±4.3	34.1±7.7	78.3±4.7	33.3±10.1
custom	CNN	G	I	10	89.3±3.9	54.9±9.0	89.4±2.2	55.4±6.2	89.6±3.1	55.9±5.8	79.0±4.3	35.4±14.0	80.5±5.0	37.6±10.4
custom	CNN	G	I	20	92.3±2.7	65.4±8.3	91.7±2.6	62.7±8.5	92.1±2.4	62.1±7.6	84.9±5.0	44.3±11.4	83.1±4.9	41.8±11.2
custom	CNN	Δ	I	3	90.4±2.5	56.8±10.5	89.0±3.3	53.5±11.6	89.0±3.1	53.7±8.6	82.2±4.4	38.6±10.9	82.1±4.4	37.8±8.8
custom	CNN	Δ	I	5	90.0±4.6	61.0±12.3	90.4±3.0	58.8±8.5	89.6±3.4	57.9±10.5	82.1±4.2	40.3±5.0	82.5±4.5	40.1±5.8
custom	CNN	Δ	I	10	92.3±2.8	62.5±11.4	89.7±4.6	59.6±8.9	89.9±4.8	58.8±10.2	83.3±4.6	43.3±8.6	83.0±4.9	40.2±10.8
custom	CNN	Δ	I	20	91.0±2.6	59.9±6.5	91.3±2.4	58.2±6.5	91.2±2.6	59.8±8.4	81.9±4.1	38.6±9.7	81.6±4.5	37.4±9.6
custom	CNN	—	—	1	90.8±3.7	59.6±8.3	91.2±2.7	59.6±8.5	91.2±2.9	59.7±7.3	82.7±4.8	39.6±9.3	82.8±5.2	40.9±7.9
Low-resolution time-of-flight mass spectrometer Pegasus® 4D-C ($\Delta = \mu_{low} - \mu_{high}$; σ omitted)														
custom	CNN	G	I	3	-10.9	-22.4	-12.4	-22.9	-11.5	-22.9	-4.8	-7.4	-3.8	-4.7
custom	CNN	G	I	5	-10.5	-21.6	-9.9	-19.7	-10.9	-19.8	-3.9	-5.6	-4.4	-4.0
custom	CNN	G	I	10	-10.4	-17.7	-10.9	-19.6	-11.4	-20.8	-1.8	-1.9	-2.9	-4.4
custom	CNN	G	I	20	-12.5	-27.8	-12.2	-26.0	-11.1	-24.1	-6.9	-10.9	-4.7	-7.9
custom	CNN	Δ	I	3	-29.0	-35.1	-27.4	-35.0	-27.0	-36.2	-20.7	-23.2	-21.2	-21.6
custom	CNN	Δ	I	5	-15.3	-30.8	-17.4	-29.8	-16.5	-28.9	-10.6	-14.2	-10.9	-12.3
custom	CNN	Δ	I	10	-14.7	-28.3	-13.3	-27.9	-12.5	-25.8	-9.5	-14.7	-8.0	-11.5
custom	CNN	Δ	I	20	-12.7	-24.8	-12.0	-22.1	-12.1	-23.3	-5.4	-6.7	-3.4	-4.4
custom	CNN	—	—	1	-11.9	-22.7	-12.7	-23.4	-12.6	-22.3	-6.1	-5.7	-5.6	-6.9

Table S5 Comparison of high- and low-resolution mass spectrometry detectors. AUROC and TPR@5%FPR results are reported for five spatial-alignment strategies applied to the channel representation (Sec. 3.2 of the main paper) of 2D Gas Chromatography Mass Spectrometry measurements from Human Scent Dataset (Sec. 4 of the main paper). For the high-resolution detector, results are reported as $\mu \pm \sigma$ across 10 folds; low-resolution rows report $\Delta = \mu_{\text{low}} - \mu_{\text{high}}$ (negative values in red; σ omitted). TPR denotes TPR@5%FPR. Identity verification is performed using a custom-designed Convolutional Neural Network (CNN). Registration uses the weighted-cosine $c_{\cos}(\bar{\mathbf{u}}_c, \bar{\mathbf{s}}_c)$ and fully convolutional $\text{FCN}(\mathbf{u}, \mathbf{s})_{\cos}^+$ compound peak detectors introduced in [3], using 22 and 24 detected compounds. Gaussian κ_G and triangular κ_Δ channel-representation kernels are used with inverse-intensity I weighting, with $C \in \{3, 5, 10, 20\}$ channels. In addition, $C = 1$ denotes the single-channel representation obtained by summing intensities across all m/z values. Reproduced from the conference version [8]

Discussion.

On Pegasus® BT 4D (high-res), the best configuration achieves **92.3 ± 2.7** AUROC and **65.4 ± 8.3** TPR@5%FPR with $G, I, C=20$; $\Delta, I, C=10$ is comparable ($92.3 \pm 2.8/62.5 \pm 11.4$). Increasing C helps up to 10–20; a single summed channel ($C=1$) is already strong ($91.2 \pm 2.9/59.7 \pm 7.3$), so multi-channel yields modest gains. Registration via $\mathcal{R}^{(22)}$ or $\mathcal{R}^{(24)}$ does not improve over no registration (differences within σ), and the FCN-cosine variant is consistently lower (≈ 8 – 10 AUROC and ≈ 20 – 24 TPR points behind across C). Relative to these high-res means, Pegasus® 4D-C (low-res) shows uniformly negative deltas; for $G, I, C=20$ the drop is -12.5 AUROC / -27.8 TPR, with the largest declines at small C and Δ kernels (e.g., $C=3$: $-29.0/-35.1$).

Overall, high-resolution appears more suitable for custom CNN; however, because cohorts

are not matched, a *within-subject paired acquisition* (matched-pairs) study — measuring the same identities on both instruments under identical protocol and timing — is required to attribute differences to detector resolution rather than cohort/batch effects.

Detector resolution num.	Total	n	Number of samples per identity			
			min	max	avg	median
low	1341	116	1	76	11.6	10
high	1187	139	2	34	8.5	9

Table S6 Statistics of detector-specific subsets of Human Scent Dataset. Pegasus® 4D-C is the low-resolution detector; Pegasus® BT 4D is the high-resolution detector. Columns: dataset name, number of samples, number of identities n , and per-identity counts (min, max, mean, median). Reproduced from the conference version [8]

S1.10 Regularization and Data Augmentation

To improve generalization under limited data, we regularized the custom CNN with dropout. In the *Backbone*, dropout (0.05) was applied in the second and third stages. In the *Down-sampling Module*, dropout rates of 0.1 and 0.15 were used in the early and late convolutional blocks, respectively (see Sec. S1.6).

We applied random Gaussian noise to the input data during training across all models for identity verification. The noise was added independently to each sample by scaling normally distributed random values with the corresponding percentage of intensities. Exactly, for each encoded measurement $\mathbf{C}$, we generated noise $\epsilon \sim \mathcal{N}(0,1)$ of the same shape and scaled it element-wise as $\epsilon \cdot (\gamma_g \cdot \mathbf{C})$ where γ_g controls the relative noise level. In all our experiments, we used $\gamma_g = 0.1$.

We employed channel-wise brightness augmentation during training. For each encoded measurement $\mathbf{C}$, the brightness in every channel k was randomly scaled by a factor $\mu_{b,k} \sim U(1 - \gamma_b, 1 + \gamma_b)$ where γ_b controls the maximum relative change. In all our experiments, we fixed $\gamma_b = 0.1$.

We further applied channel-wise contrast augmentation during training. For each encoded measurement $\mathbf{C}$, the mean intensity of each channel $\bar{\mathbf{C}}_k$ was first computed as $\bar{\mathbf{C}}_k = \frac{1}{|\Omega|} \sum_{\mathbf{u} \in \Omega} \mathbf{C}_{k,\mathbf{u}}$, where Ω denotes the set of all pixels in measurement $\mathbf{C}$. The contrast in channel k was then randomly scaled by a factor $\mu_c \sim U(1 - \gamma_c, 1 + \gamma_c)$ and applied as $\mathbf{C}'_{k,\mathbf{u}} = (\mathbf{C}_{k,\mathbf{u}} - \bar{\mathbf{C}}_k)\mu_c + \bar{\mathbf{C}}_k$ where γ_c controls the strength. In all our experiments $\gamma_c = 0.1$.

S1.11 Complete Results of CNN Embedding Encoders

In this section, complete results of experiments with CNN embedding encoders are presented in Table S7.

S1.12 Alternative Correlation-Based Representations

In this section, we investigate whether identity information can be captured from relations between spectral responses rather than from their absolute intensities.

All correlation-based representations are computed from $\log(1+x)$ -transformed intensities using Pearson correlation, implemented as the dot product of mean-centred, ℓ_2 -normalized profiles; constant or empty profiles yield zero correlations, while the diagonal of the spatial m/z representation stores normalized total m/z abundance.

“CNN”, “ResNet”, and “Small CNN” denote custom convolutional encoders of different capacities. CNN uses three to five Conv–BN–ReLU stages, with residual blocks where applicable, followed by concatenated global mean and maximum pooling. ResNet uses four residual stages. Small CNN uses a 1×1 input projection to 16 channels, followed by three stride-2, 3×3 convolutional blocks with 32, 64, and 64 channels. All encoders produce a 256-dimensional embedding and are randomly initialized and trained from scratch.

All encoders are trained using cosine-softmax identity classification with scale 24 and label smoothing 0.1, and verification uses cosine similarity between normalized embeddings; all projections are fitted separately within each outer-training fold, excluding outer-test identities and, for identity-aware projections, inner-validation identities.

S1.12.1 Gram-Based Representations

We first represent each measurement by a correlation Gram matrix and use the resulting matrix as input to a CNN classifier. We consider two complementary constructions that capture relations along either the temporal or the m/z dimension.

In the *Temporal Gram* representation, the raw binned mass spectra are grouped into P predefined temporal blocks; we use $P = 256$ in all experiments. The spectra within each block are aggregated into a single m/z spectrum, and pairwise correlations are computed between all block spectra. This yields a $P \times P$ matrix whose entry (i, j) measures the similarity between the spectral compositions of temporal blocks i and j .

Conversely, the *m/z Gram* representation captures relations between individual m/z bins. Each of the 1174 bins is represented by its intensity profile across the measurement, and pairwise correlations between these profiles produce a 1174×1174 matrix. We consider profiles constructed either from the 256 block-averaged spectra or directly from all available spectra.

			$c_{\cos}(\bar{\mathbf{u}}_c, \bar{\mathbf{s}}_c)$						FCN($\mathbf{u}, \mathbf{s}$) $^+_{\cos}$					
			no registration		$\mathcal{R}^{(22)}$		$\mathcal{R}^{(24)}$		$\mathcal{R}^{(22)}$		$\mathcal{R}^{(24)}$			
κ	w	C	AUROC	TPR	AUROC	TPR	AUROC	TPR	AUROC	TPR	AUROC	TPR		
custom CNN	G	U	3	88.3 $\pm$ 3.8	52.4 $\pm$ 9.3	87.2 $\pm$ 4.2	47.1 $\pm$ 8.0	87.3 $\pm$ 3.5	48.6 $\pm$ 7.1	84.6 $\pm$ 4.5	42.9 $\pm$ 6.8	83.4 $\pm$ 4.5	40.5 $\pm$ 7.8	
custom CNN	G	U	5	89.0 $\pm$ 2.6	50.4 $\pm$ 8.7	88.0 $\pm$ 3.5	50.7 $\pm$ 7.9	87.2 $\pm$ 3.2	50.2 $\pm$ 6.7	85.8 $\pm$ 3.4	46.5 $\pm$ 5.5	84.4 $\pm$ 3.7	41.1 $\pm$ 6.2	
custom CNN	G	U	10	88.7 $\pm$ 3.9	53.7 $\pm$ 9.7	86.2 $\pm$ 4.2	46.0 $\pm$ 10.0	85.8 $\pm$ 4.2	43.6 $\pm$ 9.2	82.2 $\pm$ 5.2	38.2 $\pm$ 8.6	83.2 $\pm$ 5.1	39.2 $\pm$ 8.5	
custom CNN	G	U	20	89.0 $\pm$ 3.7	53.1 $\pm$ 7.1	87.5 $\pm$ 3.9	49.0 $\pm$ 6.6	87.9 $\pm$ 3.0	50.3 $\pm$ 6.6	84.9 $\pm$ 3.9	43.6 $\pm$ 5.6	84.2 $\pm$ 4.8	43.4 $\pm$ 7.1	
GoogLeNet	G	U	3	87.1 $\pm$ 3.2	43.9 $\pm$ 6.9	86.5 $\pm$ 4.0	44.4 $\pm$ 7.7	86.6 $\pm$ 3.7	43.8 $\pm$ 7.7	84.1 $\pm$ 4.4	34.9 $\pm$ 8.6	84.8 $\pm$ 4.3	35.6 $\pm$ 9.4	
Inception-v3	G	U	3	86.0 $\pm$ 3.5	40.2 $\pm$ 5.6	85.5 $\pm$ 3.3	39.2 $\pm$ 5.0	85.5 $\pm$ 3.5	40.2 $\pm$ 5.1	83.5 $\pm$ 3.9	32.8 $\pm$ 6.8	84.0 $\pm$ 4.0	34.3 $\pm$ 9.1	
ResNet-101	G	U	3	85.5 $\pm$ 3.2	39.1 $\pm$ 8.2	85.4 $\pm$ 3.3	37.0 $\pm$ 4.6	85.5 $\pm$ 3.5	37.5 $\pm$ 6.4	84.3 $\pm$ 4.0	33.1 $\pm$ 6.6	84.7 $\pm$ 4.4	33.5 $\pm$ 8.7	
ResNet-18	G	U	3	87.4 $\pm$ 3.2	43.3 $\pm$ 6.5	86.6 $\pm$ 3.1	40.6 $\pm$ 5.9	86.9 $\pm$ 3.0	41.1 $\pm$ 5.9	84.7 $\pm$ 4.0	33.6 $\pm$ 7.3	85.2 $\pm$ 4.1	33.5 $\pm$ 7.7	
VGG16	G	U	3	75.0 $\pm$ 7.6	46.6 $\pm$ 12.2	78.3 $\pm$ 9.8	41.9 $\pm$ 9.4	78.4 $\pm$ 9.9	42.2 $\pm$ 8.9	73.8 $\pm$ 7.4	27.6 $\pm$ 11.2	73.8 $\pm$ 7.7	26.6 $\pm$ 8.2	
custom CNN	G	I	3	84.2 $\pm$ 7.8	44.6 $\pm$ 13.7	88.2 $\pm$ 2.9	50.8 $\pm$ 5.1	86.7 $\pm$ 3.9	46.8 $\pm$ 8.0	84.0 $\pm$ 4.5	40.5 $\pm$ 5.6	83.8 $\pm$ 5.1	40.0 $\pm$ 7.8	
custom CNN	G	I	5	89.5 $\pm$ 3.4	53.3 $\pm$ 8.3	88.3 $\pm$ 3.4	51.3 $\pm$ 6.5	88.4 $\pm$ 3.4	50.7 $\pm$ 6.7	85.6 $\pm$ 4.4	43.5 $\pm$ 7.5	85.4 $\pm$ 3.8	43.9 $\pm$ 6.2	
custom CNN	G	I	10	83.6 $\pm$ 5.2	38.9 $\pm$ 12.2	87.0 $\pm$ 4.1	43.2 $\pm$ 8.4	84.5 $\pm$ 5.2	41.6 $\pm$ 8.7	83.0 $\pm$ 4.7	36.8 $\pm$ 9.8	83.3 $\pm$ 4.9	39.5 $\pm$ 8.3	
custom CNN	G	I	20	84.7 $\pm$ 4.1	48.1 $\pm$ 3.4	87.1 $\pm$ 2.9	49.5 $\pm$ 5.8	86.1 $\pm$ 4.7	48.1 $\pm$ 5.5	88.5 $\pm$ 1.1	53.6 $\pm$ 2.3	87.0 $\pm$ 4.1	46.2 $\pm$ 1.7	
GoogLeNet	G	I	3	87.9 $\pm$ 3.5	46.8 $\pm$ 7.7	86.8 $\pm$ 3.9	44.3 $\pm$ 8.4	86.7 $\pm$ 3.4	44.7 $\pm$ 7.5	85.0 $\pm$ 3.8	35.2 $\pm$ 8.9	84.9 $\pm$ 4.1	36.1 $\pm$ 10.3	
Inception-v3	G	I	3	85.4 $\pm$ 3.5	43.0 $\pm$ 9.0	84.3 $\pm$ 3.6	39.8 $\pm$ 5.6	84.5 $\pm$ 3.7	40.6 $\pm$ 6.2	82.6 $\pm$ 3.9	32.4 $\pm$ 9.1	82.8 $\pm$ 3.7	33.6 $\pm$ 10.8	
ResNet-101	G	I	3	86.3 $\pm$ 4.1	41.5 $\pm$ 11.4	86.0 $\pm$ 4.2	38.5 $\pm$ 8.7	86.5 $\pm$ 4.1	40.6 $\pm$ 7.7	84.5 $\pm$ 4.7	35.3 $\pm$ 8.7	84.9 $\pm$ 4.4	35.2 $\pm$ 9.1	
ResNet-18	G	I	3	87.5 $\pm$ 3.9	44.9 $\pm$ 8.8	87.3 $\pm$ 3.5	42.9 $\pm$ 5.2	86.9 $\pm$ 3.2	41.9 $\pm$ 5.9	85.2 $\pm$ 3.8	36.1 $\pm$ 7.1	85.5 $\pm$ 4.3	36.2 $\pm$ 8.2	
VGG16	G	I	3	76.5 $\pm$ 7.3	47.3 $\pm$ 12.9	79.5 $\pm$ 9.6	42.9 $\pm$ 10.4	79.8 $\pm$ 9.4	42.2 $\pm$ 9.4	74.6 $\pm$ 7.2	26.2 $\pm$ 8.3	74.5 $\pm$ 6.9	26.6 $\pm$ 6.6	
custom CNN	Δ	U	3	55.4 $\pm$ 4.9	3.4 $\pm$ 4.2	55.6 $\pm$ 5.3	3.3 $\pm$ 4.2	55.7 $\pm$ 4.5	3.8 $\pm$ 3.6	55.5 $\pm$ 5.0	3.3 $\pm$ 3.5	55.6 $\pm$ 4.8	3.5 $\pm$ 3.9	
custom CNN	Δ	U	5	55.5 $\pm$ 4.6	3.4 $\pm$ 4.3	55.6 $\pm$ 5.5	3.5 $\pm$ 4.3	55.4 $\pm$ 4.7	3.4 $\pm$ 3.8	54.9 $\pm$ 4.4	3.6 $\pm$ 3.7	55.0 $\pm$ 3.8	3.3 $\pm$ 3.8	
custom CNN	Δ	U	10	55.4 $\pm$ 4.7	3.7 $\pm$ 4.1	55.3 $\pm$ 4.9	3.6 $\pm$ 4.2	54.8 $\pm$ 4.3	3.2 $\pm$ 3.9	55.4 $\pm$ 4.2	3.9 $\pm$ 4.5	55.7 $\pm$ 5.1	3.2 $\pm$ 3.7	
custom CNN	Δ	U	20	55.8 $\pm$ 5.7	3.6 $\pm$ 4.5	55.1 $\pm$ 4.2	3.7 $\pm$ 3.8	55.7 $\pm$ 5.1	3.8 $\pm$ 4.7	54.8 $\pm$ 4.0	3.4 $\pm$ 4.2	55.0 $\pm$ 4.2	3.4 $\pm$ 3.8	
GoogLeNet	Δ	U	3	53.0 $\pm$ 3.0	1.2 $\pm$ 1.7	52.9 $\pm$ 2.7	1.2 $\pm$ 1.4	53.2 $\pm$ 3.3	1.7 $\pm$ 2.0	52.9 $\pm$ 3.0	1.3 $\pm$ 1.6	53.0 $\pm$ 3.1	1.3 $\pm$ 1.6	
Inception-v3	Δ	U	3	56.0 $\pm$ 5.4	3.4 $\pm$ 4.2	55.9 $\pm$ 5.3	3.3 $\pm$ 4.2	56.0 $\pm$ 5.1	3.6 $\pm$ 3.8	55.6 $\pm$ 5.0	3.2 $\pm$ 3.6	55.9 $\pm$ 5.3	3.4 $\pm$ 4.0	
ResNet-101	Δ	U	3	52.8 $\pm$ 2.5	1.2 $\pm$ 1.7	52.6 $\pm$ 2.3	1.2 $\pm$ 1.4	52.6 $\pm$ 2.5	1.7 $\pm$ 2.0	52.4 $\pm$ 2.1	1.3 $\pm$ 1.6	56.1 $\pm$ 11.1	5.0 $\pm$ 11.5	
ResNet-18	Δ	U	3	51.9 $\pm$ 1.9	0.7 $\pm$ 0.8	51.9 $\pm$ 1.8	0.8 $\pm$ 0.9	52.1 $\pm$ 2.4	1.2 $\pm$ 1.8	51.9 $\pm$ 1.9	0.9 $\pm$ 0.9	52.1 $\pm$ 2.4	0.9 $\pm$ 1.2	
VGG16	Δ	U	3	52.4 $\pm$ 2.3	1.3 $\pm$ 1.7	52.5 $\pm$ 2.3	1.2 $\pm$ 1.4	52.6 $\pm$ 2.9	1.9 $\pm$ 1.9	52.4 $\pm$ 2.4	1.5 $\pm$ 1.5	52.7 $\pm$ 2.8	1.4 $\pm$ 1.5	
custom CNN	Δ	I	3	83.1 $\pm$ 8.4	42.9 $\pm$ 13.7	81.0 $\pm$ 5.1	35.2 $\pm$ 6.8	80.4 $\pm$ 4.7	36.2 $\pm$ 5.0	78.7 $\pm$ 5.9	31.8 $\pm$ 6.6	77.8 $\pm$ 6.1	29.8 $\pm$ 7.6	
custom CNN	Δ	I	5	82.3 $\pm$ 7.0	43.5 $\pm$ 5.1	85.5 $\pm$ 4.1	43.5 $\pm$ 7.4	85.4 $\pm$ 3.6	43.9 $\pm$ 6.8	80.6 $\pm$ 5.7	34.6 $\pm$ 9.2	77.8 $\pm$ 6.3	27.6 $\pm$ 9.7	
custom CNN	Δ	I	10	87.1 $\pm$ 1.3	51.3 $\pm$ 4.8	86.8 $\pm$ 3.7	52.4 $\pm$ 5.9	86.3 $\pm$ 3.9	50.0 $\pm$ 6.8	82.7 $\pm$ 4.9	42.2 $\pm$ 8.4	81.4 $\pm$ 5.0	37.0 $\pm$ 9.6	
custom CNN	Δ	I	20	82.3 $\pm$ 5.2	37.4 $\pm$ 10.5	81.5 $\pm$ 3.1	51.3 $\pm$ 4.2	87.3 $\pm$ 2.5	48.1 $\pm$ 5.1	88.1 $\pm$ 1.1	51.2 $\pm$ 3.5	86.4 $\pm$ 2.4	49.1 $\pm$ 3.7	
GoogLeNet	Δ	I	3	84.0 $\pm$ 3.3	36.4 $\pm$ 8.0	84.1 $\pm$ 3.8	32.2 $\pm$ 6.1	83.9 $\pm$ 3.6	31.5 $\pm$ 6.0	82.0 $\pm$ 3.2	25.3 $\pm$ 7.1	81.7 $\pm$ 3.1	23.3 $\pm$ 5.3	
Inception-v3	Δ	I	3	83.4 $\pm$ 3.4	35.7 $\pm$ 8.4	82.5 $\pm$ 3.8	28.7 $\pm$ 10.3	83.0 $\pm$ 3.5	28.4 $\pm$ 8.5	79.6 $\pm$ 3.6	24.1 $\pm$ 9.1	79.7 $\pm$ 3.3	22.7 $\pm$ 8.6	
ResNet-101	Δ	I	3	86.0 $\pm$ 3.4	40.4 $\pm$ 9.7	83.6 $\pm$ 2.9	32.5 $\pm$ 4.8	83.5 $\pm$ 3.3	32.6 $\pm$ 6.0	81.7 $\pm$ 1.8	27.2 $\pm$ 4.1	81.4 $\pm$ 1.9	27.7 $\pm$ 4.1	
ResNet-18	Δ	I	3	83.5 $\pm$ 4.1	33.3 $\pm$ 9.8	83.0 $\pm$ 4.3	26.2 $\pm$ 9.4	83.0 $\pm$ 3.9	26.4 $\pm$ 9.5	80.6 $\pm$ 3.9	21.4 $\pm$ 7.6	81.0 $\pm$ 3.9	21.2 $\pm$ 7.9	
VGG16	Δ	I	3	77.0 $\pm$ 4.6	40.5 $\pm$ 6.6	77.4 $\pm$ 3.4	38.1 $\pm$ 5.9	77.4 $\pm$ 3.4	38.0 $\pm$ 6.0	71.7 $\pm$ 2.7	25.1 $\pm$ 3.4	70.8 $\pm$ 2.8	25.2 $\pm$ 3.6	

Table S7 All cross-validation AUROC and TPR@5%FPR results for five spatial-alignment strategies on the channel representation (Sec. 3.2 of the main paper) of 2D Gas Chromatography Mass Spectrometry measurements on Human Scent Dataset (Sec. 4 of the main paper). Results are reported as $\mu \pm \sigma$ across 10 folds. TPR denotes TPR@5%FPR. Identity verification is performed with: (i) a custom-designed Convolutional Neural Network (CNN), and (ii) five pretrained neural-network backbones: *ResNet-18* [2], *ResNet-101* [2], *VGG-16* [7], *Inception-v3* [9], and *GoogLeNet* [9], each with an appended and trained projection layer. Registration is performed with the weighted-cosine $c_{\cos}(\bar{\mathbf{u}}_c, \bar{\mathbf{s}}_c)$ and fully convolutional FCN($\mathbf{u}, \mathbf{s}$) $^+_{\cos}$ compound peak detectors introduced in [3], using 22 and 24 detected compounds. Gaussian κ_G and triangular κ_Δ channel-representation kernels are evaluated with uniform U and inverse-intensity I weighting, with $C \in \{3, 5, 10, 20\}$ channels. Reproduced from the conference version [8]

All variants are evaluated on the full Human Scent Dataset using the same 10-fold identity-disjoint cross-validation protocol as the main method. For the Temporal Gram representation, we evaluate both our lightweight CNN and a ResNet backbone, while the higher-dimensional m/z Gram matrices are evaluated using ResNet.

The Gram-based representations retain substantial identity information (Tab. S8) but consistently underperform compared to the proposed representations, suggesting that absolute spectral responses contain complementary discriminative information that is lost when measurements are reduced to pairwise correlations.

S1.12.2 Spatial Correlation Representations

We next investigate correlation-based representations that explicitly preserve the two-dimensional chromatographic layout. The motivation is to test whether identity information is encoded not only in the spectra themselves, but also in how spectral responses are related across the chromatographic plane. To reduce the computational cost, we first apply spatial binning by aggregating neighboring raw pixels into 10×10 blocks. We consider two complementary constructions.

The first representation is a t_1 -column correlation tensor. After spatial binning, each binned pixel is represented by its aggregated m/z spectrum. For each pair of binned t_1 columns, we compute Pearson correlations between every pair of their t_2 -position spectra, producing a full $t_2 \times t_2$ correlation matrix. Stacking these matrices over all t_1 -column pairs yields the final tensor.

As a complementary variant, we construct a *spatial m/z correlation* representation. Here, each m/z channel is treated as a two-dimensional intensity map over the chromatographic plane. These maps are vectorized over spatial positions, and pairwise correlations are computed between all m/z channels. The resulting representation captures which mass channels exhibit similar spatial distributions, thereby testing whether identity information is reflected in their spatial co-occurrence patterns rather than in their absolute intensities.

Both variants are evaluated on the full Human Scent Dataset using the same 10-fold identity-disjoint cross-validation protocol as the main method. The t_1 -column correlation representation retains a substantial amount of identity information, whereas correlations between spatial m/z distributions are considerably less discriminative (Tab. S8).

S1.12.3 Learned projection for raw-pixel aggregation

We hypothesized that spatial averaging used in previous representation might be losing some information. Instead of replacing each raw block by a uniform average, we learned separable row and column projections inside each block. Therefore, simple averaging is just a degenerated case of this method.

Representation	Backbone	AUROC	TPR@5%
<i>Gram-based</i>			
Temporal Gram	CNN	84.6 ± 4.2	36.5 ± 10.0
Temporal Gram	ResNet	85.9 ± 4.7	40.4 ± 9.8
m/z Gram (blocks)	ResNet	84.6 ± 6.4	38.9 ± 15.3
m/z Gram (all)	ResNet	80.5 ± 6.0	29.4 ± 8.3
<i>Spatial correlations</i>			
t_1 correlation	CNN	87.3 ± 3.7	43.0 ± 11.1
Spatial m/z correlation	CNN	80.2 ± 4.1	25.6 ± 6.5
<i>Learned projection</i>			
Global	CNN	85.3 ± 4.6	32.8 ± 8.3
Global	Small CNN	82.8 ± 7.8	32.4 ± 7.3
Block-specific	CNN	83.3 ± 6.4	34.4 ± 7.9
Block-specific	Small CNN	86.5 ± 3.3	33.7 ± 6.8

Table S8 Performance of alternative correlation-based representations on the full Human Scent Dataset. Results are mean $\pm$ standard deviation over 10 identity-disjoint folds. All values are in percent

We first trained the projection using an unsupervised objective based on maximizing the entropy of the resulting representation:

$$\mathcal{L}_{\text{ent}} = -\frac{1}{2} \log \det \left(\text{Cov}(Z) + \epsilon \frac{\text{tr}(\text{Cov}(Z))}{K} I \right) + \lambda [D_{\text{KL}}(a||u_R) + D_{\text{KL}}(b||u_C)].$$

Here, K is the dimensionality of the projected representation, a and b are the normalized row- and column-projection weights, and u_R and u_C are the corresponding uniform distributions.

We computed covariance either across blocks within a measurement or across measurements at the same position. Neither produced a stable non-trivial solution: weak regularization selected a single sub-pixel, whereas strong regularization approached uniform averaging.

We therefore also tested an identity-aware projection trained to increase between-identity variability while reducing within-identity variability. The learned weights were then frozen and used to construct the same t_1 -correlation representation. We evaluated both a global separable projection shared across all spatial blocks and block-specific row and column projections. Although the weights did not collapse, verification performance remained below that of the main multi-view representation (Tab. S8).

References

- [1] Cuzuel V, Leconte R, Cognon G, et al (2018) Human odor and forensics: Towards Bayesian suspect identification using GC×GC–MS characterization of hand odor. *Journal of Chromatography B* 1092:379–385. <https://doi.org/10.1016/j.jchromb.2018.06.018>
- [2] He K, Zhang X, Ren S, et al (2016) Deep Residual Learning for Image Recognition. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp 770–778, <https://doi.org/10.1109/CVPR.2016.90>
- [3] Hlavsa J, Spetlik R, Čechová J, et al (2025) Sex Classification from Human Scent Using Image Interpretation of 2D Gas Chromatography-Mass Spectrometry Data. In: Petersen J, Dahl VA (eds) *Image Analysis*. Springer Nature Switzerland, pp 457–470, https://doi.org/10.1007/978-3-031-95911-0_32
- [4] Ojala T, Pietikainen M, Maenpaa T (2002) Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on pattern analysis and machine intelligence* 24(7):971–987
- [5] Pojmanová P, Ladislavová N, Škeříková V, et al (2020) Human scent samples for chemical analysis. *Chemical Papers* 74(5):1383–1393. <https://doi.org/10.1007/s11696-019-00989-2>
- [6] Pojmanová P, Ladislavová N, Urban Š (2021) Development of a Method for the Measurement of Human Scent Samples Using Comprehensive Two-Dimensional Gas Chromatography with Mass Detection. *Separations* 8(12):232. <https://doi.org/10.3390/separations8120232>
- [7] Simonyan K, Zisserman A (2015) Very Deep Convolutional Networks for Large-Scale Image Recognition. In: *Proceedings of 3rd International Conference on Learning Representations*
- [8] Spetlik R, Hlavsa J, Čechová J, et al (2026) Identity verification from human scent using channel representation of 2d gas chromatography-mass spectrometry data. In: *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pp 1471–1480
- [9] Szegedy C, Liu W, Jia Y, et al (2015) Going deeper with convolutions. In: *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pp 1–9, <https://doi.org/10.1109/CVPR.2015.7298594>