Technical Report: One-Class Surface Deformation Forecasting for FaceMoCap (v2)

1 Scope and objective

This report documents the methodology implemented in `oneclass_surface_deformation_v2.py` for one-class anomaly detection on FaceMoCap 3D facial motion capture sequences. The pipeline is *movement-specific*: for each facial movement $m \in \{1, \ldots, 5\}$, it (i) selects a region of interest (ROI) on the face using healthy subjects only, (ii) extracts deformation features from a fixed facial topology, (iii) trains a temporal forecaster on healthy data only, and (iv) scores sequences using forecast errors to detect deviations from healthy dynamics.

2 Data model

2.1 Raw marker sequences

Each recording is a time-indexed sequence of 3D points. After reading a sample CSV, the data are reshaped per frame into

$$\mathbf{P}_t \in \mathbb{R}^{108 \times 3}, \quad t = 1, \ldots, T,$$

where points $0, 1, 2$ are head reference markers (dental support) and points $3..107$ are facial markers. The facial subset is

$$\mathbf{Q}_t \in \mathbb{R}^{105 \times 3}, \quad \mathbf{Q}_t = \mathbf{P}_t[3 : 108, :].$$

Missing markers are represented as NaNs and are handled via binary masks.

2.2 Facial topology

A fixed template connectivity is provided as a triangular mesh over the 105 facial markers:

$$\text{faces} = \{(a_k, b_k, c_k)\}_{k=1}^{K_f}, \quad a_k, b_k, c_k \in \{0, \ldots, 104\}.$$

The undirected edge set induced by triangles is

$$E = \text{unique}\Big(\big\{ \{a_k, b_k\}, \{b_k, c_k\}, \{c_k, a_k\} \; : \; k = 1, \ldots, K_f \big\} \Big). \tag{1}$$

3 Head-frame normalization

The pipeline removes global head motion by mapping facial points into a per-frame head coordinate system derived from the three head reference markers. Let $\mathbf{p}_{t,0}, \mathbf{p}_{t,1}, \mathbf{p}_{t,2} \in \mathbb{R}^3$ be the head markers in frame t. Define orthonormal axes

$$\mathbf{x}_t = \frac{\mathbf{p}_{t,1} - \mathbf{p}_{t,0}}{\|\mathbf{p}_{t,1} - \mathbf{p}_{t,0}\|_2}, \tag{2}$$

$$\tilde{\mathbf{z}}_t = \mathbf{x}_t \times (\mathbf{p}_{t,2} - \mathbf{p}_{t,0}), \qquad \mathbf{z}_t = \frac{\tilde{\mathbf{z}}_t}{\|\tilde{\mathbf{z}}_t\|_2}, \tag{3}$$

$$\mathbf{y}_t = \mathbf{z}_t \times \mathbf{x}_t, \tag{4}$$

and the rotation matrix

$$\mathbf{R}_t = [\mathbf{x}_t \ \mathbf{y}_t \ \mathbf{z}_t] \in \mathbb{R}^{3 \times 3}. \tag{5}$$

A facial point $\mathbf{q}_{t,j}$ in world coordinates is mapped to head coordinates via

$$\mathbf{q}_{t,j}^{(\text{head})} = \mathbf{R}_t^\top (\mathbf{q}_{t,j} - \mathbf{p}_{t,0}). \tag{6}$$

If the head frame is degenerate (e.g., collinear markers or missing values), the transform for that frame is invalid and the corresponding facial points are treated as missing.

4 ROI marker selection from healthy data

ROI selection is performed independently for each movement using healthy subjects only.

4.1 Per-sample peak displacement

For a given sample and marker j, define a rest pose as the mean of the first K valid frames (default: $K = \texttt{rest\_first\_valid\_frames}$):

$$\mathbf{r}_j = \frac{1}{|\mathcal{T}_{j,K}|} \sum_{t \in \mathcal{T}_{j,K}} \mathbf{q}_{t,j}^{(\text{head})}, \tag{7}$$

where $\mathcal{T}_{j,K}$ are the first K time indices where marker j is valid. Define the displacement magnitude

$$d_{t,j} = \left\| \mathbf{q}_{t,j}^{(\text{head})} - \mathbf{r}_j \right\|_2, \tag{8}$$

and a robust peak displacement (default: 95th percentile)

$$p_j = \text{percentile}_q \big(\{ d_{t,j} : t \in \mathcal{T}_j \} \big), \qquad q = 0.95, \tag{9}$$

where $\mathcal{T}_j$ are all valid times for marker j in the sample.

4.2 Movement-specific z-scores

For each movement m, let $\mu_m(j)$ be the mean of p_j across healthy samples of movement m. Let $\mu_g(j)$ and $\sigma_g(j)$ be the mean and standard deviation of $\mu_m(j)$ across movements (global statistics over movements). A z-score is computed as

$$z(m,j) = \frac{\mu_m(j) - \mu_g(j)}{\sigma_g(j) + \varepsilon}. \tag{10}$$

The ROI marker set S_m is defined as the top-K markers ranked by $z(m,j)$ (default: $K = \texttt{K\_markers}$).

5 ROI edges from topology

Given the global edge set E from Eq. (1), ROI edges for movement m are those fully contained in S_m:

$$E_m = \{ (u,v) \in E : u \in S_m \wedge v \in S_m \}. \tag{11}$$

If $|E_m|$ is below a minimum threshold (default: $\texttt{min\_roi\_edges}$), the ROI marker set is expanded by including mesh neighbors within a fixed number of hops ($\texttt{expand\_hops}$), then Eq. (11) is recomputed.

6 Deformation feature extraction

6.1 Neutral frame selection

For each sample, a neutral frame t_0 is selected from the first 10% of frames as the frame with:
(i) a valid head transform and (ii) the maximum number of valid facial markers (ties broken by
earliest time).

6.2 Edge-length deformation

For each ROI edge $e = (i, j) \in E_m$, define the neutral length

$$d_0(e) = \left\| \mathbf{q}_{t_0,i}^{(\text{head})} - \mathbf{q}_{t_0,j}^{(\text{head})} \right\|_2, \tag{12}$$

and the length at time t

$$d_t(e) = \left\| \mathbf{q}_{t,i}^{(\text{head})} - \mathbf{q}_{t,j}^{(\text{head})} \right\|_2.$$

Two feature definitions are supported:

$$x_t(e) = d_t(e) - d_0(e) \qquad (\text{absolute}), \tag{13}$$

$$x_t(e) = \frac{d_t(e) - d_0(e)}{d_0(e) + \varepsilon} \qquad (\text{relative}). \tag{14}$$

A binary feature mask indicates validity:

$$m_t(e) = \mathbb{1}\left[\text{endpoints } i, j \text{ valid at } t \text{ and at } t_0\right]. \tag{15}$$

Stacking features over ROI edges yields a multivariate time series

$$\mathbf{x}_t \in \mathbb{R}^D, \quad D = |E_m|, \qquad \mathbf{m}_t \in \{0,1\}^D.$$

7 Active-window cropping

A per-frame motion energy signal is computed from available features:

$$\text{energy}(t) = \frac{\sum_{e \in E_m} |x_t(e)| \, m_t(e)}{\max\left(1, \sum_{e \in E_m} m_t(e)\right)}. \tag{16}$$

Define the active index set

$$\mathcal{A} = \{t : \text{energy}(t) \geq \alpha \cdot \max_u \text{energy}(u)\}, \tag{17}$$

with $\alpha = \texttt{win\_thr\_frac}$. An active window $[a, b]$ is formed around $\mathcal{A}$ with padding ($\texttt{win\_pad}$)
and minimum length ($\texttt{win\_min\_len}$). If $\mathcal{A}$ is empty, a peak-centered fallback window is used.

8 Resampling with missingness constraints

The cropped sequence is resampled to a fixed length T_{out} using linear interpolation applied
feature-wise, while avoiding interpolation across large gaps. Let $\{t_k\}$ be original time indices
where feature d is valid ($m_{t_k,d} = 1$). For a target time τ, the interpolated value uses the nearest
bracketing valid times $t_k \leq \tau \leq t_{k+1}$. The value is accepted only if

$$(t_{k+1} - t_k) \leq G_{\max}, \tag{18}$$

where $G_{\max} = \texttt{max\_gap}$ (in original-frame units); otherwise the resampled mask at τ is set to 0
and the value is treated as missing.

This yields resampled arrays

$$\mathbf{X} \in \mathbb{R}^{T_{\text{out}} \times D}, \quad \mathbf{M} \in \{0,1\}^{T_{\text{out}} \times D}.$$

9 Robust normalization (healthy train only)

For each movement, a robust scaler is fit on *healthy training* samples only, ignoring masked entries. For feature dimension d:

$$\text{med}_d = \text{nanmedian}(X_{\cdot,d}), \qquad \text{IQR}_d = Q_{75}(X_{\cdot,d}) - Q_{25}(X_{\cdot,d}). \tag{19}$$

Normalization is

$$\tilde{X}_{t,d} = \frac{X_{t,d} - \text{med}_d}{\text{IQR}_d + \varepsilon} M_{t,d}. \tag{20}$$

The multiplication by $M_{t,d}$ ensures missing entries remain at 0 after normalization.

10 One-class temporal forecasting model

A separate temporal model is trained for each movement. The input at each time step concatenates values and masks:

$$\mathbf{u}_t = [\tilde{\mathbf{X}}_t \parallel \mathbf{M}_t] \in \mathbb{R}^{2D}.$$

The implemented forecaster is a masked Temporal Convolutional Network (TCN): a 1×1 projection to a hidden size followed by residual dilated 1D convolutions over time, and an output projection producing $\hat{\mathbf{X}}_t \in \mathbb{R}^D$.

10.1 Masked Huber training loss

Training uses a next-step prediction style (time indices $t = 2, \ldots, T_{\text{out}}$), with a masked Huber loss. For residual r, the Huber penalty is

$$\ell_\delta(r) = \begin{cases} \frac{1}{2}r^2, & |r| \leq \delta, \\ \delta\left(|r| - \frac{1}{2}\delta\right), & |r| > \delta, \end{cases} \tag{21}$$

with $\delta = \texttt{huber\_delta}$. The masked sequence loss is

$$\mathcal{L} = \frac{\sum_{t=2}^{T_{\text{out}}} \sum_{d=1}^{D} \ell_\delta\big(\hat{X}_{t,d} - X_{t,d}\big) M_{t,d}}{\max\left(1, \sum_{t=2}^{T_{\text{out}}} \sum_{d=1}^{D} M_{t,d}\right)}. \tag{22}$$

11 Anomaly scoring

Given predicted features $\hat{\mathbf{X}}$, a masked per-time error is computed (MAE):

$$e_t = \frac{\sum_{d=1}^{D} |\hat{X}_{t,d} - X_{t,d}| M_{t,d}}{\max\left(1, \sum_{d=1}^{D} M_{t,d}\right)}, \qquad t = 2, \ldots, T_{\text{out}}. \tag{23}$$

A weight w_t is derived from the resampled motion energy and per-frame availability (implementation: energy scaled and then normalized to $[0, 1]$ after down-weighting by availability). The main score is the weighted mean error

$$s = \frac{\sum_{t=2}^{T_{\text{out}}} w_t e_t}{\max\left(\varepsilon, \sum_{t=2}^{T_{\text{out}}} w_t\right)}. \tag{24}$$

A secondary score may be used as a worst-case indicator:

$$s_{\max} = \max_t e_t.$$

12   Threshold selection and evaluation protocol

Splits are computed at the *participant level* using healthy subjects only (train/validation/test, default: 0.70/0.15/0.15). The anomaly threshold for movement m is set as a percentile of validation healthy scores:

$$\tau_m = \text{percentile}_p\big(\{s_i : i \in \text{validation healthy}\}\big), \qquad p = \texttt{thr\_percentile}. \tag{25}$$

A sample is flagged anomalous if $s > \tau_m$. Pathological samples are not used for training or thresholding; they are used only for scoring and reporting.

13   Outputs

A run produces a timestamped output directory containing: (i) dataset audit tables and plots, (ii) ROI marker sets and ROI edge lists per movement, (iii) preprocessing sanity artifacts, (iv) per-movement trained models and scaler statistics, (v) per-movement score CSVs, histograms, and threshold summaries, and (vi) a global summary table and run configuration dump.

14   Implementation notes and failure modes

- **Head-frame degeneracy:** if head reference markers are missing or nearly collinear, Eq. (5)–(6) is invalid for that frame; all derived features for that frame become missing.

- **Gap-aware interpolation:** Eq. (18) prevents bridging long missing segments, but can increase missingness after resampling on low-quality sequences.

- **ROI sensitivity:** ROI markers are selected statistically from healthy movement data (Eq. (10)); if healthy training data are heterogeneous, ROI definitions may broaden.

- **One-class assumption:** thresholds (Eq. (25)) reflect healthy variability; they do not guarantee optimal separation for any particular pathology subtype.