Compute heatmap embeddings

Radim Nedbal

Function space generated by the composition of embedding, wavelet band reconstruction, and nonlinear ICA

Let $X \in \mathbb{R}^{I\times J\times T}$ denote a time series of facial heatmaps with spatial size $I \times J$ and temporal length T. We consider three operators applied in cascade:

- **Spatial embedding (framewise)**
  $\mathcal{E} : \mathbb{R}^{I\times J} \to \mathbb{R}^K$, applied to each frame x_t, producing $\mathcal{E}(X) \in \mathbb{R}^{K\times T}$.
- **Per-channel 1D DWT and band reconstruction**
  $\Omega : \mathbb{R}^{K\times T} \to \mathbb{R}^{K\times T}$, applying a 1D discrete wavelet transform to each channel, selecting bands, and reconstructing.
- **Nonlinear ICA / dimensionality reduction**
  $\Psi : \mathbb{R}^{K\times T} \to \mathbb{R}^{L\times T}$ with $L < K$, applied framewise or on short windows.

Define the composed map

$$f = \Psi \circ \Omega \circ \mathcal{E} \colon \mathbb{R}^{I\times J\times T} \to \mathbb{R}^{L\times T} \ .$$

The function space is

$$\mathcal{F} = \{\, f : \mathbb{R}^{I\times J\times T} \to \mathbb{R}^{L\times T} \mid f = \Psi \circ \Omega \circ \mathcal{E} \text{ for admissible } \mathcal{E}, \Omega, \Psi \,\}.$$

Typical assumptions

1. **Regularity of $\mathcal{E}$**: linear or smooth (PCA, moments, or differentiable neural embedding).

2. **Wavelet properties**: Ω denotes the 1D discrete wavelet analysis followed by optional subband selection and synthesis. Selecting or zeroing coefficients (i.e., discarding subbands) is a projection in coefficient space and in general destroys invertibility; the reconstructed signal is then the band-limited approximation obtained from the retained subbands.
3. **Nonlinear ICA identifiability**: Ψ trained with auxiliary/temporal structure enabling identifiability.

Analytic and geometric properties

- **Finite parameterization:** $\mathcal{F}$ is a finite-dimensional manifold when operators are finitely parameterized:

Let $\Theta \subset \mathbb{R}^p$ parameterize the stages $\mathcal{E}_{\theta_E}, \Omega_{\theta_\Omega}, \Psi_{\theta_\Psi}$ and define

$$\Phi : \Theta \to \mathcal{M}, \qquad \Phi(\theta) = \Psi_{\theta_\Psi} \circ \Omega_{\theta_\Omega} \circ \mathcal{E}_{\theta_E},$$

where $\mathcal{M}$ is a Banach space of maps (for example C^0 or L^2 mappings). The induced function class is the image $\Phi(\Theta) = \{f_\theta : \theta \in \Theta\} \subset \mathcal{M}$. If Φ is C^r and an immersion at θ, then $\Phi(\Theta)$ is a p-dimensional C^r submanifold of $\mathcal{M}$. Parameter redundancies reduce the effective dimension, and nonparametric or infinite-width stages may produce infinite-dimensional families.

- **Band-limited structure:** outputs after Ω are localized in time–frequency.

> **Note**
>
> Time-localization is useful because it lets you detect, separate, and interpret short, emotion-related transients (micro-events) from background noise; it improves denoising, feature extraction, and interpretability for time-varying facial heatmaps, especially when emotions produce brief thermal signatures.

- **Dimensionality reduction:** effective dimension reduces from $K \times T$ to $L \times T$.

We employ a sequencewise encoder

$$\Psi : \mathbb{R}^{K\times T} \to \mathbb{R}^{L\times T} \quad (\text{or } \Psi : \mathbb{R}^{K\times T} \to \mathbb{R}^{L\times T'} \text{ if temporal downsampling is used}),$$

so that the effective spatio-temporal degrees of freedom are reduced from $K \times T$ to $L \times T$ (or $L \times T'$). This formulation lets Ψ exploit temporal dependencies for identifiability and for isolating emotion-relevant transients while compressing redundant or noisy dimensions.

- **Identifiability constraints:** independence assumptions carve out identifiable latent subspaces.

> **Note**
>
> Independence (or conditional-independence) assumptions are the mechanism that restrict the space of possible latent explanations and thereby make certain nonlinear-ICA problems identifiable: they "carve out" a subspace (or *quotient*) of latent solutions that can be uniquely recovered up to trivial indeterminacies.