Multimodal Autoencoder for BP, Respiration, and PPG

Md Farhan Tasnim Oshim

1 Goal

Reconstruct 8-second windows (240 samples @ 30 Hz) of three vital biosignals:`bp`, `breath_upper_upper`, `ppg_fing`.

Preprocessing & ingestion

- Parsed CSVs, verified required columns.
- **Downsampled** to 30 Hz with FIR decimation (zero-phase).
- **Windowed** into 8-s segments with 50% overlap.
- **Per-modality standardization** using `StandardScaler`, fit on valid (non-corrupted) segments.
- **Corruption handling:** any all-NaN or flatline channel is masked & replaced with zeros; masks carried through losses.

Auto-encoder Architecture

- One **Conv1d encoder per modality** → latent codes.
- **Fusion MLP** (concat latents → linear → tanh → dropout=0.5) to get shared representation.
- One **decoder per modality** (MLP → upsample → Conv1d) mapping fused latent back.

Training objective

- **Time-domain loss:** weighted masked MSE.
- **Frequency-domain loss:** masked spectral MSE on normalized rFFT magnitudes.
- **Total loss:** $\alpha \cdot L_{time} + \beta \cdot L_{freq}$ (final runs: $\alpha = 0.5, \beta = 0.5$).
- Early stopping on validation loss.
- Noise injection integrated in training loop for denoising experiments.

2 Assumptions

- Signal content within an 8-s window is locally stationary enough for spectral comparison.
- Zero-phase FIR decimation suffices for antialiasing across subjects.
- Per-modality scaling is appropriate despite cross-subject amplitude differences.
- Corrupted segments are not learnable targets; masking avoids biasing the loss.
- Spectral loss on magnitude only (phase-agnostic) is acceptable for morphology-driven signals.

3 Quantitative Results

Representative results from one run are shown in Table 1.

Table 1: Quantitative performance summary.

Category	Metric	Value / Notes
Optimization	Best Val Loss (total)	0.689 (Time 1.378 + 0.5 × Freq 0.000639)
Optimization	Test Loss (total)	0.441
Global Reconst	MSE (time domain)	0.398
Global Reconst	MAE (time domain)	0.398
Global Reconst	SNR (dB)	7.09 dB
Global Reconst	Spectral Coherence	∼1.0*
Compression	Input size (floats)	720
Compression	Bottleneck size (floats)	128
Compression	Compression ratio	5.625×
Denoising	ΔSNR (bp/breath_upper/ppg_fing)	-15.88 / -20.77 / -20.98 dB
Cross-Modal	MAE (bp/breath_upper/ppg_fing)	9.81 / 8.87 / 0.77
Cross-Modal	SNR (dB) (bp/breath_upper/ppg_fing)	17.75 / -2.39 / 0.34
Cross-Modal	Spectral Cosine (bp/breath_upper/ppg_fing)	0.998 / 0.949 / 0.642

*Spectral score near 1.0 reflects strong DC; detrending before spectral metrics yields more discriminative values.

4 Qualitative Results

The qualitative comparison of the reconstructed signals (both time and frequency domains) can be found in the ipynb file. The reconstructions for `bp` and `ppg_fing` demonstrate reasonable fidelity to the original signals, capturing essential waveform and spectral characteristics. However, the reconstruction quality for `breath_upper` is noticeably poorer, likely due to a higher proportion of missing or corrupted data in that modality, which impedes accurate modeling.

5 What worked

- Adding spectral term stabilized training and preserved dominant rhythms.
- Masked losses prevented corrupted channels from contaminating optimization.
- Per-modality encoders + shared latent + decoders handled bandwidths reasonably well.

6 What didn't

- Denoising failed ($\Delta\text{SNR} < 0$) since the model was not explicitly trained as a denoiser.
- Spectral similarity is overly forgiving in the presence of large DC.
- Loss curves suggest overfitting occurs (despite early stopping, weight decay, dropout, etc.) because of a small dataset (10 CSV files) to generalize; combined with possible modality imbalance and very limited validation data, the model easily memorizes training patterns.

7 Future Work

In the future, the following areas could be explored to enhance the model:

1. **Advanced Architectures:** While the 1D-CNN performed well, exploring UNet, recurrent architectures like LSTMs, or attention-based models like Transformers could potentially improve the capture of long-range temporal dependencies across the 8-second windows.

2. **Data Augmentation:** Given the limited dataset of 10 subjects, implementing time-series specific data augmentation techniques (e.g., time warping or magnitude scaling) could improve the model's robustness and generalization.

3. **Hyperparameter Tuning:** A more exhaustive search could be performed to tune hyperparameters such as the learning rate, the latent dimension size, and the alpha and beta weights in the composite loss function to find an optimal balance between time and frequency domain reconstruction.