Multimodal A2: Sensor Fusion with Attention

Rajeev Atla

1. Introduction

Human activity recognition underpins many wellness applications, from rehabilitation and fall detection to everyday fitness tracking. Inertial measurement units are a natural starting point for motion understanding, but they struggle during low-motion or stationary states such as sleep or quiet rest. Physiological signals like heart rate supply complementary context that can stabilize predictions in those edge cases. Fusing physiological and kinematic data can therefore lift robustness and overall accuracy, especially when sensors are noisy or degraded.

This report studies multimodal fusion on the PAMAP2 Physical Activity Monitoring dataset [1], [2], which aligns heart rate with IMU streams from the accelerometer, gyroscope, and magnetometer across 18 activities. The task is challenging for three reasons: different sampling rates between modalities, missing or unreliable measurements that include IMU drift and noisy heart rate, and temporal offsets such as delayed heart rate changes after a subject sits down. Fusion models also risk overfitting to the denser modality, which hurts validation performance and transfer when inputs go missing.

2. Approach

We compare three fusion strategies for activity recognition: early, late, and hybrid fusion.

Early fusion: Signals are temporally aligned and concatenated at the input level, then processed by a shared encoder. This captures joint correlations but can overfit to the most informative or highest-rate stream and absorb its noise.

Late fusion: Each modality is encoded by a separate transformer, and predictions are combined near the output. This preserves modularity and can improve calibration, but it may miss fine-grained cross-modal interactions that arise earlier in the pipeline.

Hybrid fusion: Separate encoders are linked with cross-attention so that features exchange information at multiple depths. Each encoder continues to learn modality-specific structure while cross-attention adjusts the relative influence of signals in context.

3. Results

3.1. Fusion Comparison

3.2. Attention

3.3. Uncertainty Calibration

3.4. Ablation Studies

4. Discussion

5. Conclusion

6. References

[1] A. Reiss and D. Stricker, "Introducing a New Benchmarked Dataset for Activity Monitoring," in *2012 16th International Symposium on Wearable Computers*, 2012, pp. 108–109. doi: 10.1109/ISWC.2012.13.

[2] A. Reiss, "PAMAP2 Physical Activity Monitoring." 2012.