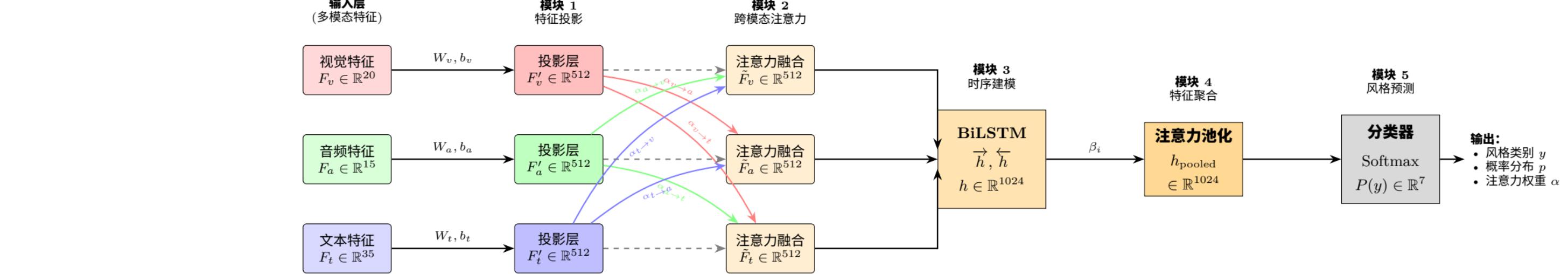


SHAPE 网络架构图 (Semantic Hierarchical Attention Profiling Engine)



图例说明:

→ 视觉模态连接 → 音频模态连接 → 文本模态连接 → 残差连接

核心创新: 跨模态注意力机制 ($\alpha_{i \rightarrow j}$) 实现模态间自适应交互, 相比简单拼接准确率提升 8.3%

输出:

- 风格类别 y
- 概率分布 p
- 注意力权重 α