\***拟投稿分论坛：材料大数据与智能技术**

**基于特征增强的材料成分多模态融合网络**

**于烨泳1，武星1,2,3,4，钱权1,2,3,4[[1]](#footnote-1)\***

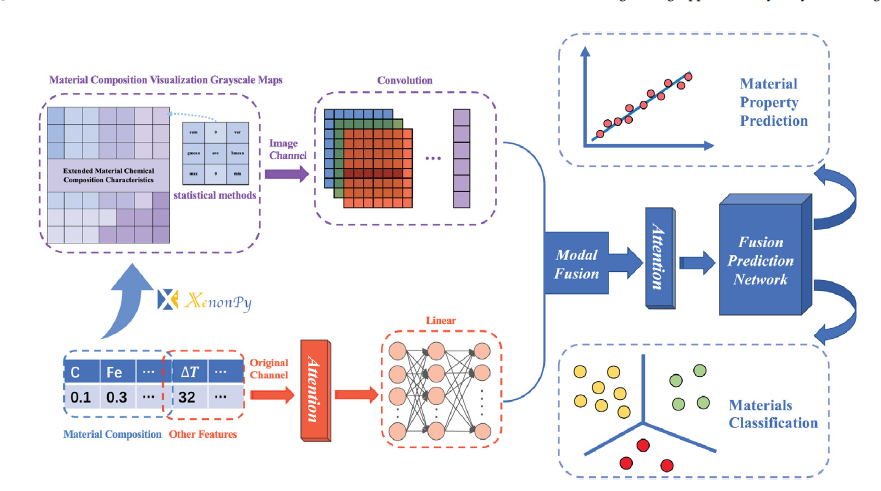
1 上海大学，计算机工程与科学学院，上海，200444，中国；

2 上海大学，材料基因组工程研究院，上海，200444，中国；

3 上海大学，先进通讯和数据科学研究院，上海，200444，中国；

4 之江实验室，杭州，211100，中国；

**详细摘要：**鉴于先进高性能材料的复杂性和多样性，全面了解材料的成分-过程-结构-性能关系具有挑战性。 数据驱动方法已被视为新材料研发的第四范式。 然而，许多材料数据集中组分元素的复杂性导致了非常稀疏的组成特征，给机器学习模型带来了巨大的挑战。在本研究中，应用基于原子特征扩充并通过统计公式进行数据增强的方案，将材料成分特征转换为元素级别统计特征，随后利用相关性排列将增强特征进行模态转换，以此来缓解材料小数据集中维度灾难的问题。本研究将化学成分特征映射为二维灰度图像数据，以解决稀疏材料成分矩阵的问题，基于此，提出了一种材料组成可视化网络（MCVN），根据灰度特征映射方案，单模态数据集扩充为多模态。然后将包含在图像的模态特征中的隐含信息映射为灰度图中的纹理，通过CNN进行高效地滤波噪声并提取关键信息；接着通过多头自注意力模型和全连接层进行提取，最后进行模态融合输出材料的性能和分类结果；并应用于预测钢的力学性能和非晶合金材料分类。我们将MCVN与其他机器学习方法进行了比较。 在国家材料科学研究所钢材数据集的四个目标上，MCVN的平均𝑅2值提高了4％，而其他模型已经达到了平均𝑅2值为0.92；在上海材料研究所钢的横截面收缩目标数据集中， 它达到了0.835的𝑅2值，而其他模型的平均𝑅2值仅为0.64。 对于不平衡的非晶合金材料数据集，MCVN将小类晶体合金（CRA）的平均𝑅𝑒𝑐𝑎𝑙𝑙从0.58提高到0.78。 基于扩展材料化学成分信息的方法具有通用性，为材料性能预测提供了新的范式。



**图1**. 材料成分增强与模态融合网络

**关键词：** 材料成分特征密集化；模态转换；多模态融合；



**参考文献：**[1] Yu Y, Wu X, Qian Q. Better utilization of materials’ compositions for predicting their properties: Material composition visualization network[J]. Engineering Applications of Artificial Intelligence, 2023, 117: 105539.

**报告人简介：**

钱权教授，毕业于中国科学技术大学计算机应用专业，目前任上海大学计算机学院智能科学系主任。他长期从事材料大数据、机器学习及工业化应用、网络安全、隐私计算等研究，主持完成国家重点研发计划及省部级科研项目30余项。已发表学术论文100余篇，拥有发明专利和软件著作权50余项。

1. \* 通讯作者: qqian@shu.edu.cn、13917469688. [↑](#footnote-ref-1)