Densification and Visualization of Material Composition Characteristics for Enhanced Materials Property Prediction

Yeyong Yu [a], Xing Wu [a], Quan Qian [a, b, *]

a. School of Computer Engineering and Science, Shanghai University

b. Materials Genome Institute, Shanghai University, Shanghai

7th Forum of Materials Genome Engineering

Introduction

- Complex material datasets challenge machine learning due to sparse features.
- Composition crucial: composition proportions overlooked.
- Compositional features present: Atomic info missing.
- XenonPy: 58 elemental points, 406 extended dimensions.
- Grayscale images for CNN extraction.

Methods

Densification and Prediction: Atomic features for 2D grayscale images; MCVN predicts steel properties, classifies amorphous alloys. [See Figure A]

- **Expert-based visualization**: Enhances compositional features in material ML tasks.
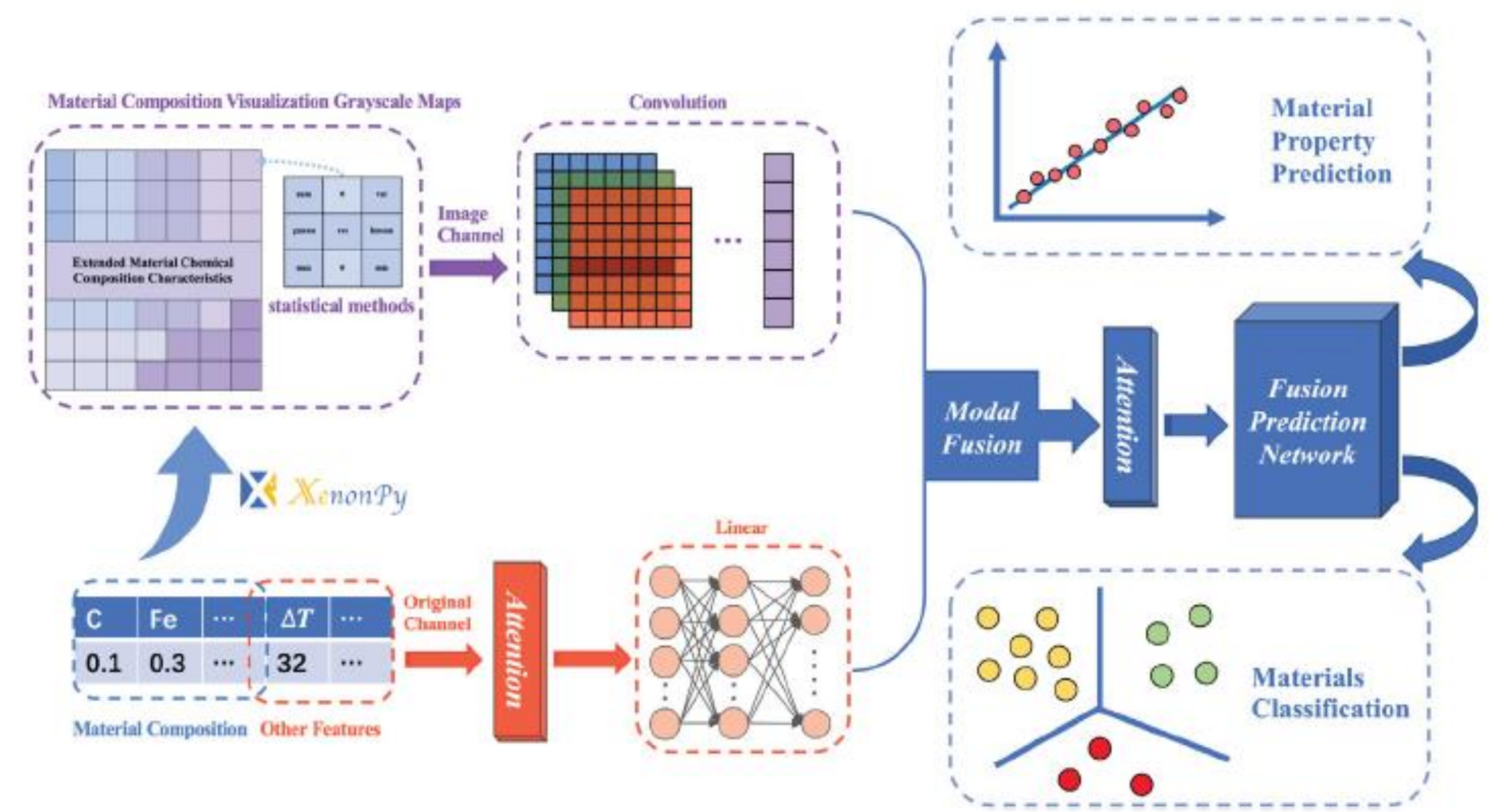- **Multimodal deep-network model**: Integrates visualization images & raw features, improves accuracy. [See Figure B]

Figure A. Better utilization of composition for material property prediction using the MCVN.
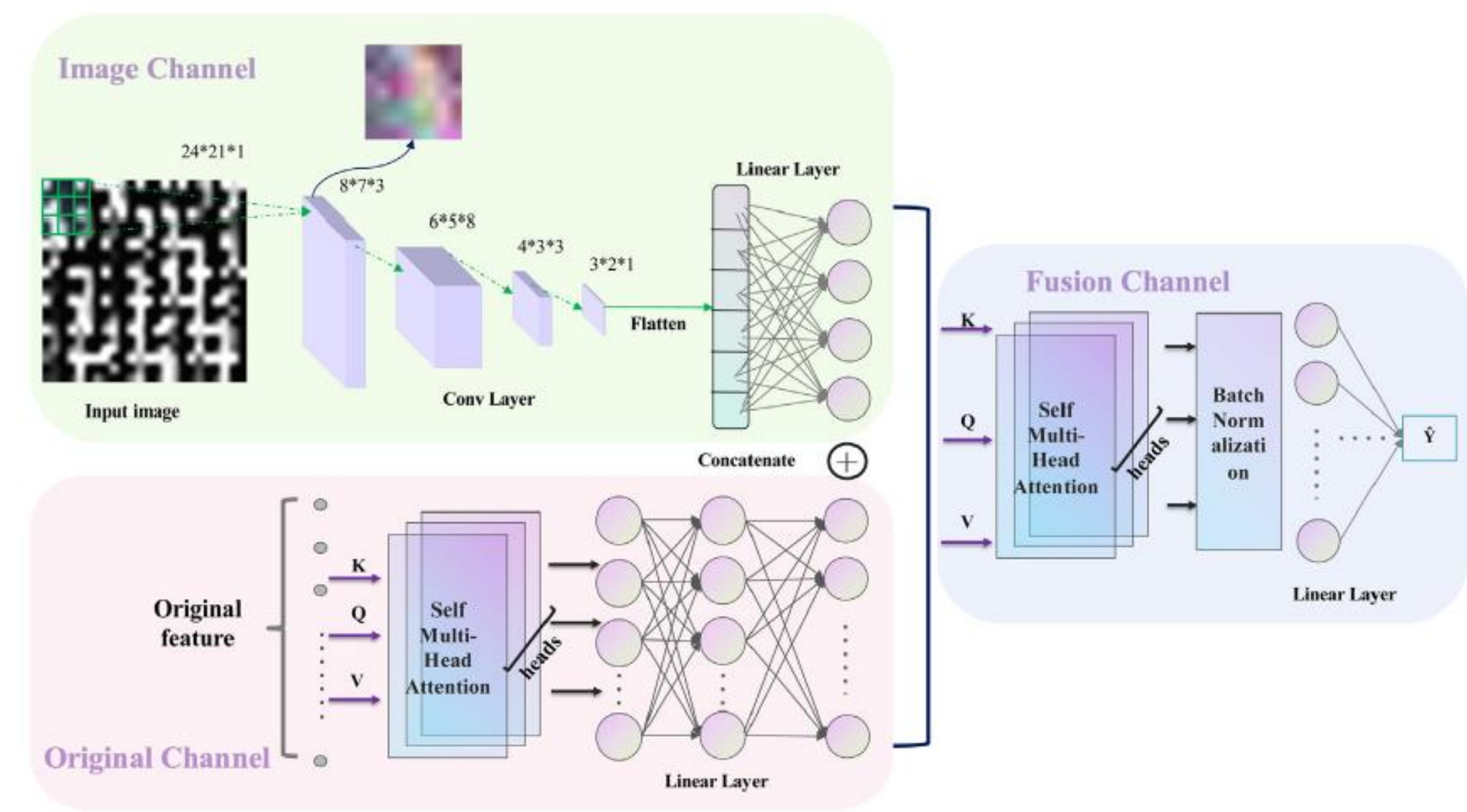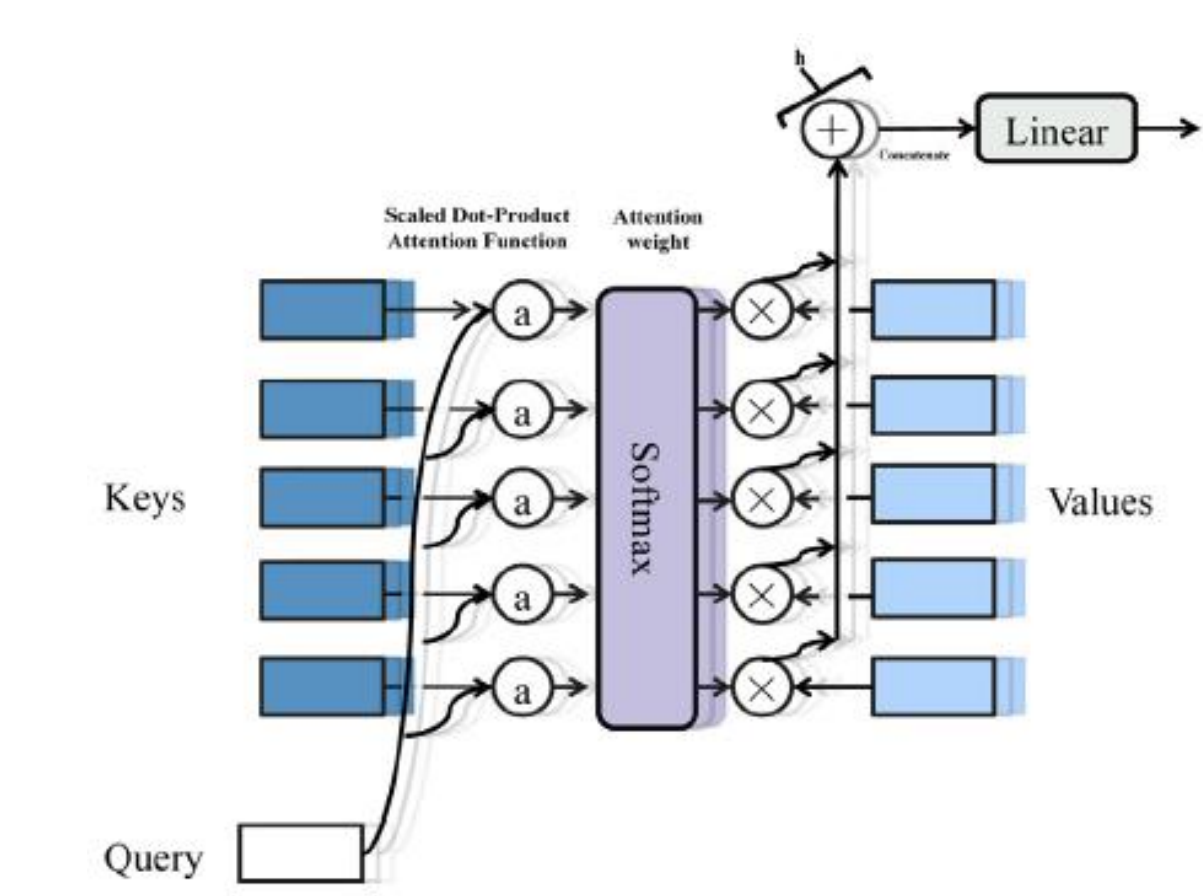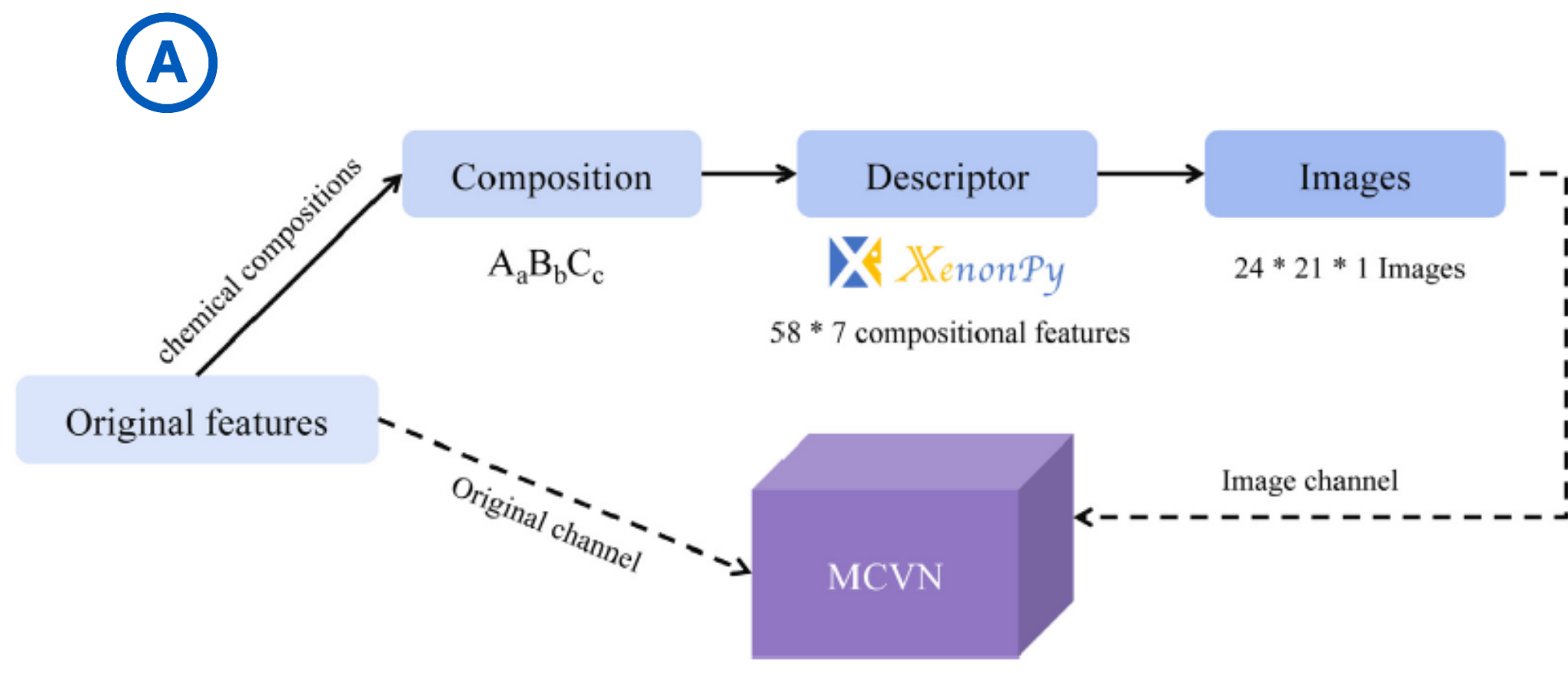

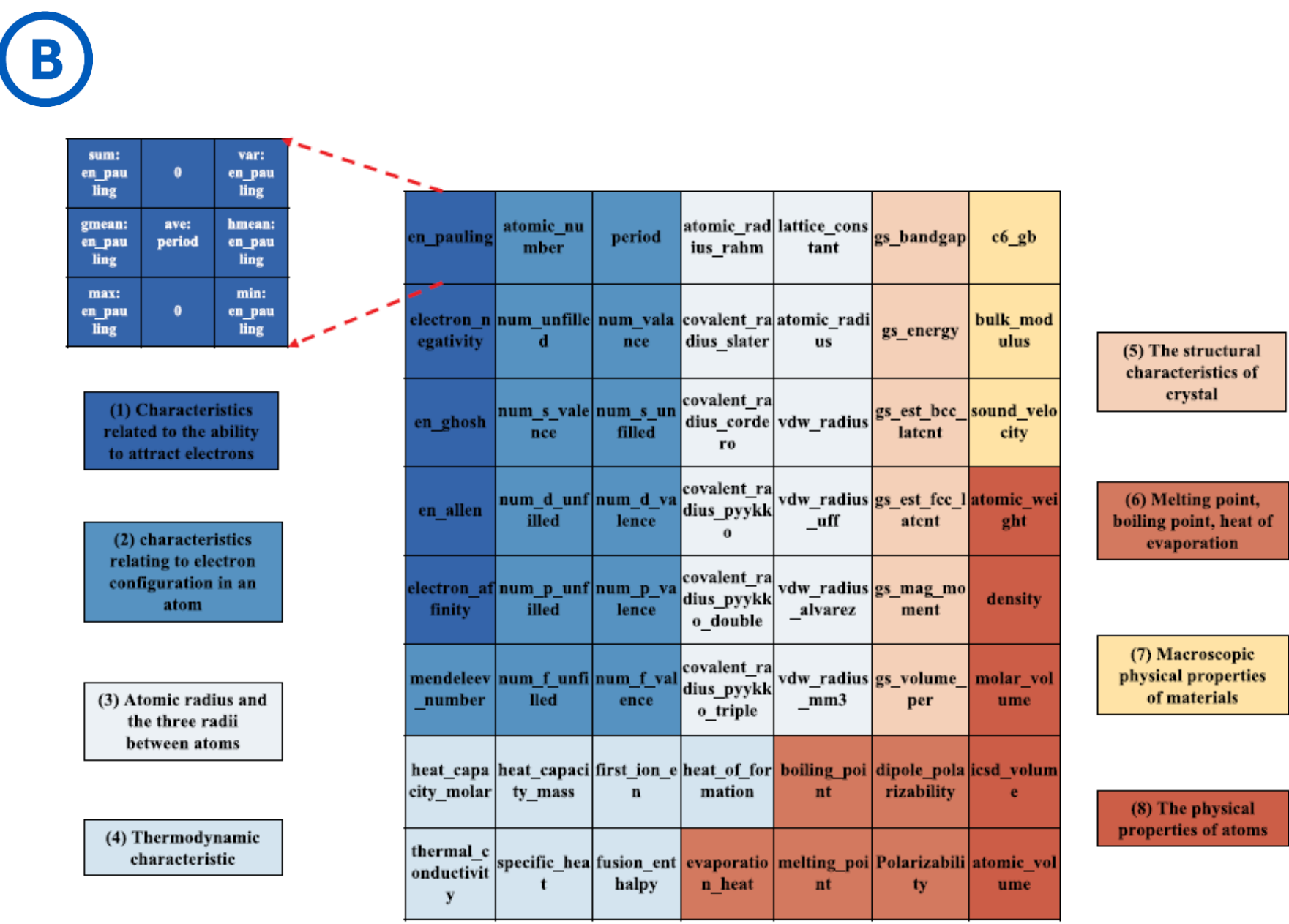**Figure B.** Material composition visualization network architecture.

Self-attention Mechanism: Reduces external info reliance, captures internal relevance effectively.
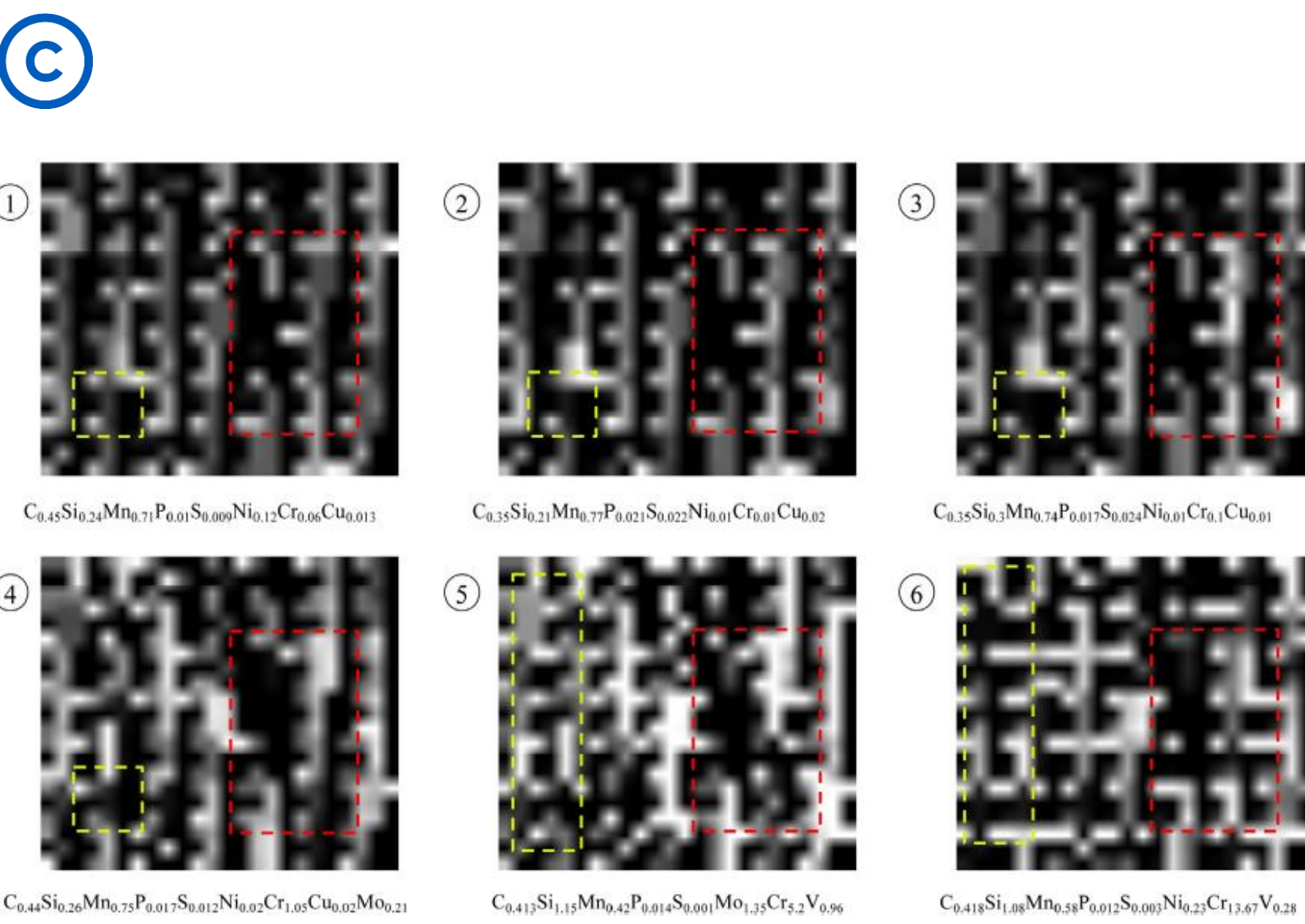
Modal Translation

(A)

Process Flow: Feature Enhancement with Xenonpy.

(B)

**Arrangement:** Template of compositional features.

(C)

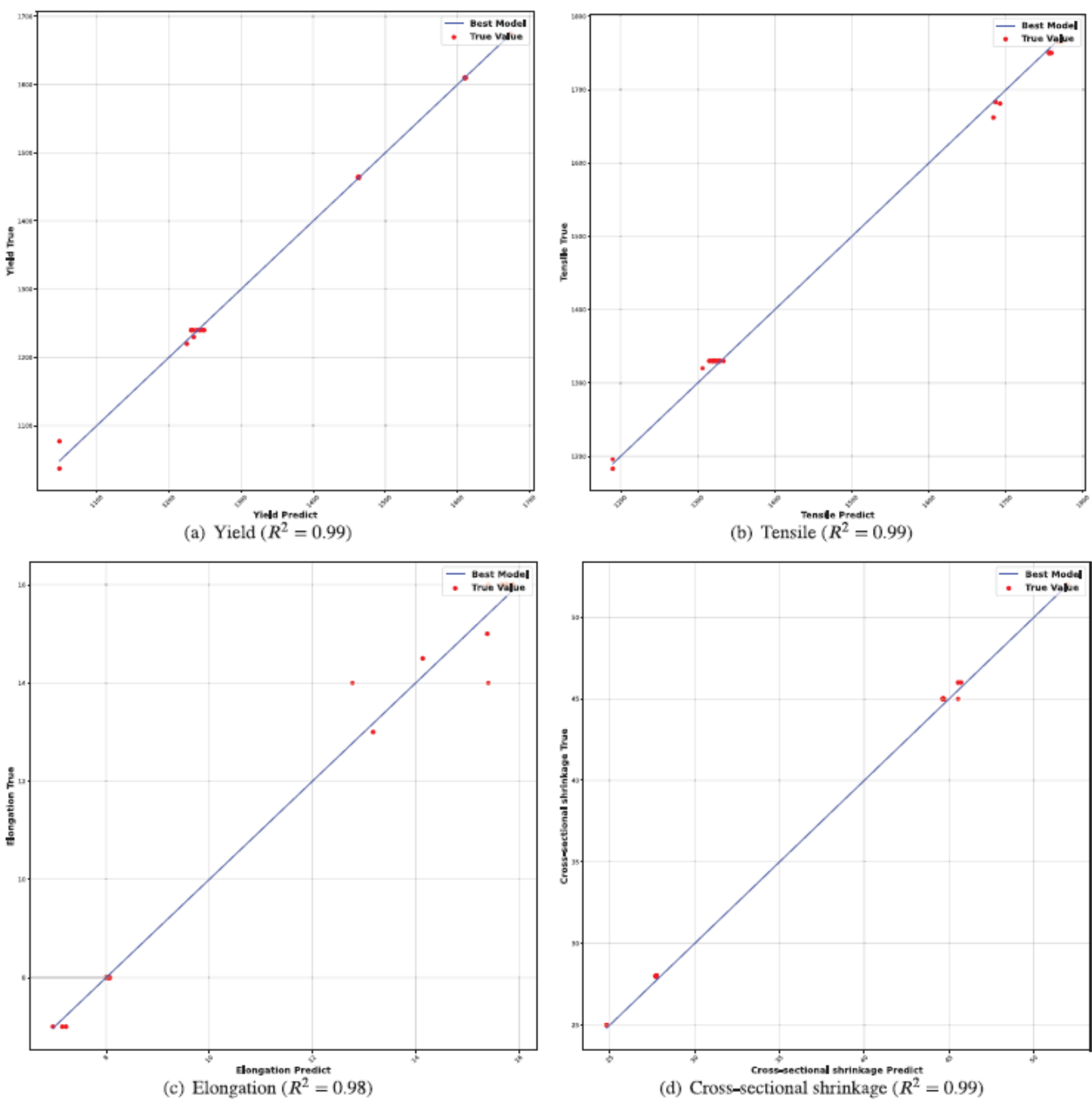**Grayscale Image:** Element-level description of features

Results

Regression Task

Target	MCVN	Attention Network (Vaswani et al., 2017)			edRVFL_N (Shi et al., 2022)			Ridge (hottou, 2010)			SVR (Chang and Lin, 2011)			RF (Breiman, 2015)			XGBoost (Chen and Guestrin, 2016)		
		O	E	E&P	O	E	E&P	O	E	E&P	O	E	E&P	O	E	E&P	O	E	E&P
Fatigue	0.954	0.937	0.926	0.939	0.915	0.949	0.913	0.909	0.913	0.906	0.89	0.780	0.882	0.903	0.886	0.904	0.920	0.931	0.921
Tensile	0.983	0.966	0.958	0.968	0.936	0.975	0.952	0.939	0.945	0.938	0.925	0.785	0.916	0.919	0.919	0.90	0.942	0.951	0.928
Fracture	0.954	0.937	0.942	0.947	0.869	0.920	0.864	0.894	0.915	0.905	0.905	0.859	0.907	0.913	0.941	0.929	0.925	0.939	0.934
Hardness	0.980	0.965	0.954	0.968	0.923	0.960	0.926	0.927	0.940	0.929	0.907	0.783	0.899	0.927	0.934	0.904	0.935	0.948	0.931

(a) Fatigue (R^2 = 0.954)

(b) Tensile (R^2 = 0.983)

(c) Fracture (R^2 = 0.954)

(d) Hardness (R^2 = 0.98)
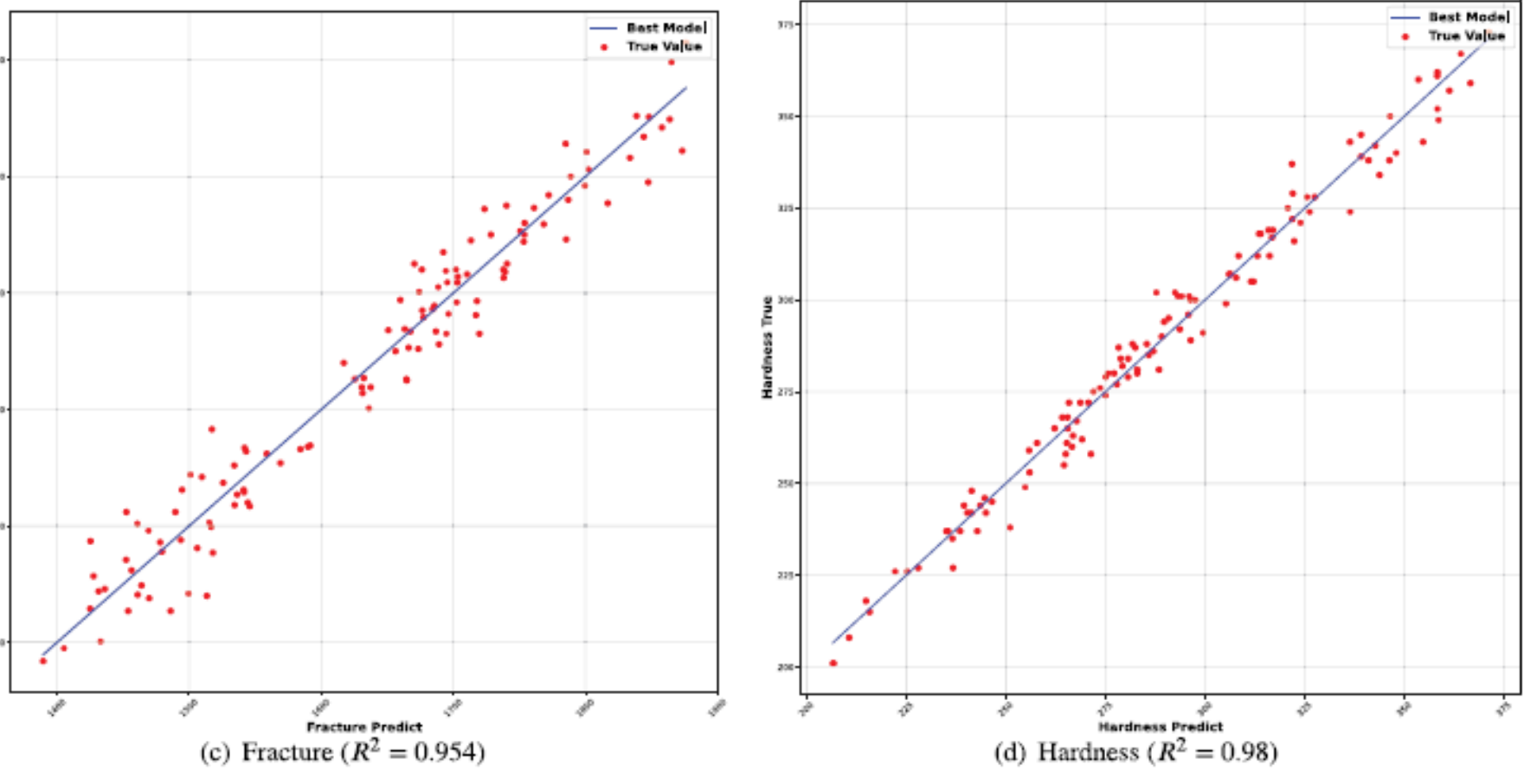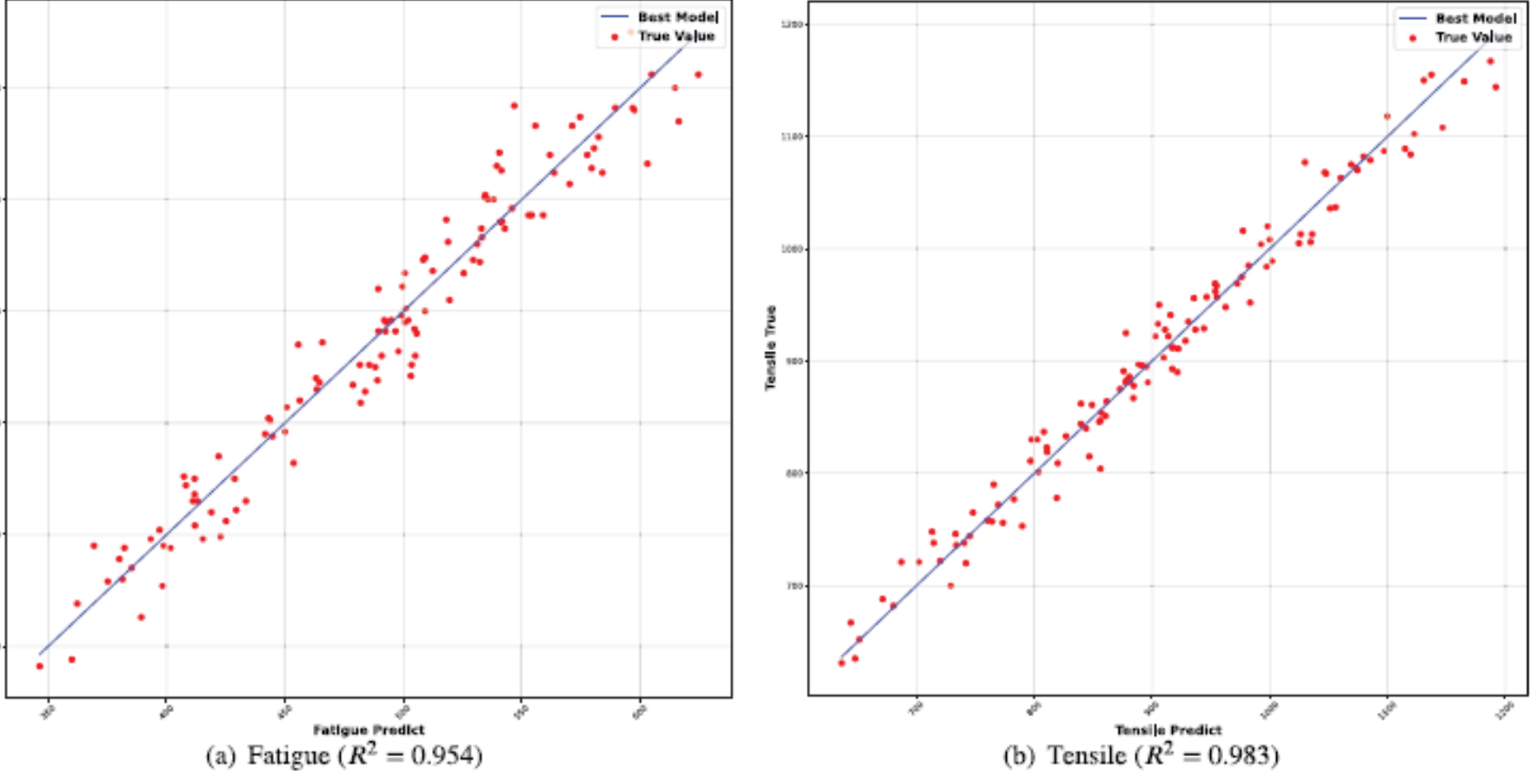
NIMS's steel dataset: 4% R^2 improvement (0.92 avg. R^2)

Regression Task

Target	MCVN	Attention Network (Vaswani et al., 2017)			edRVFL_N (Shi et al., 2022)			Ridge (hottou, 2010)			SVR (Chang and Lin, 2011)			RF (Breiman, 2015)			XGBoost (Chen and Guestrin, 2016)		
		O	E	E&P	O	E	E&P	O	E	E&P	O	E	E&P	O	E	E&P	O	E	E&P
Yield	0.974	0.932	0.912	0.938	0.928	0.947	0.940	0.923	0.949	0.928	0.891	0.899	0.880	0.926	0.947	0.944	0.908	0.926	0.922
Tensile	0.98	0.942	0.924	0.956	0.957	0.970	0.956	0.944	0.959	0.942	0.918	0.905	0.878	0.94	0.958	0.933	0.915	0.948	0.956
Elongation	0.943	0.912	0.897	0.927	0.843	0.869	0.872	0.882	0.77	0.858	0.875	0.858	0.859	0.839	0.860	0.830	0.833	0.614	0.793
Cross-sectional shrinkage	0.835	0.638	0.593	0.723	0.729	0.701	0.755	0.624	0.596	0.654	0.544	0.617	0.582	0.668	0.556	0.658	0.656	0.411	0.605

(a) Yield (R^2 = 0.99)

(b) Tensile (R^2 = 0.99)

(c) Elongation (R^2 = 0.98)

(d) Cross-sectional shrinkage (R^2 = 0.99)

SRIM's steel dataset: R^2 = 0.835 on EL (vs 0.64 avg. R^2).

Classification Task

Algorithm	Category		Precision	Recall	F1-score	Support
RF (Breiman, 2015)	BMG	O	0.83	0.94	0.87	
		E	0.72	0.93	0.82	208
		E&P	0.77	0.92	0.84	
	RMG	O	0.82	0.94	0.87	
		E	0.83	0.92	0.87	1241
		E&P	0.81	0.92	0.86	
	CRA	O	0.86	0.54	0.67	
		E	0.88	0.54	0.67	510
		E&P	0.83	0.50	0.62	
XGBoost (Chen and Guestrin, 2016)	BMG	O	0.92	0.85	0.89	
		E	0.92	0.87	0.90	208
		E&P	0.92	0.80	0.86	
	RMG	O	0.83	0.95	0.89	
		E	0.86	0.94	0.90	1241
		E&P	0.81	0.95	0.87	
	CRA	O	0.83	0.57	0.68	
		E	0.84	0.67	0.75	510
		E&P	0.83	0.54	0.66	
Attention Network (Vaswani et al., 2017)	BMG	O	0.842	0.875	0.858	
		E	0.895	0.860	0.877	208
		E&P	0.899	0.861	0.879	
	RMG	O	0.871	0.905	0.888	
		E	0.890	0.905	0.898	1241
		E&P	0.885	0.913	0.899	
	CRA	O	0.786	0.583	0.684	
		E	0.794	0.562	0.678	510
		E&P	0.806	0.596	0.70	
edRVFL_N (Shi et al., 2022)	BMG	O	0.936	0.567	0.706	
		E	0.967	0.860	0.910	208
		E&P	0.937	0.581	0.718	
	RMG	O	0.770	0.923	0.840	
		E	0.847	0.930	0.887	1241
		E&P	0.785	0.930	0.851	
	CRA	O	0.742	0.503	0.600	
		E	0.798	0.643	0.712	510
		E&P	0.777	0.547	0.642	
MCVN	BMG		0.931	0.904	0.917	208
	RMG		0.901	0.941	0.921	1241
	CRA		0.861	0.778	0.818	510

Unbalanced amorphous alloy dataset: Increased avg. *Recall* (CRA small-class) from 0.58 to 0.78.

Conclusion

This approach enhances model performances on small sparse matrix samples by expanding material composition data, offering a universal paradigm for predicting material properties.

- **Feature Enhancement:** Densifies sparse composition, boosts generalization.
- **Modal Translation:** XenonPy, element-level features, multimodal grayscale dataset.
- **Multimodal Learning:** MCVN enhances material property predictions.

Reference

1. **Better utilization of materials' compositions for predicting their properties: Material composition visualization network[J]. Engineering Applications of Artificial Intelligence**
2. **FTAP: Feature transferring autonomous machine learning pipeline[J]. Information Sciences**