

FusionX-Net: Cross-Attention Enhanced Masked Autoencoders for Multi-Modal Remote Sensing Data Fusion

Nowshad Hasan¹

Chittagong University of Engineering and Technology¹

INTRODUCTION

- **Introduction:** In remote sensing, multi-modal data fusion is crucial for combining different types of satellite data to gain richer insights.
- **Problem:** Multi-modal data fusion, particularly combining SAR (Synthetic Aperture Radar) and optical imagery, **faces challenges due to domain gaps and high labeling costs** in remote sensing.
- **Existing Solutions:** Contrastive learning has been widely used, but it **relies heavily on complex data augmentation techniques**, which require domain expertise and large compute resources.
- **Proposed Solution:** FusionX-Net offers **an alternative approach by using masked autoencoders and a cross-attention mechanism** to integrate SAR and optical data without needing extensive data augmentation.

METHOD

FusionX-Net uses masked autoencoders and cross-attention to fuse SAR and optical data:

- **Cross-Attention Encoder:** Combines data early to capture cross-modal relationships.
- **Masked Autoencoders (MAE):** Masks parts of the input data and reconstructs them to learn useful features.
- **Multi-Task Decoder:** Reconstructs both SAR and optical data with feature-level fusion.
- **Masking Strategies:**
 - **Independent:** Masks patches separately across both modalities.
 - **Consistent:** Masks the same patches for both.

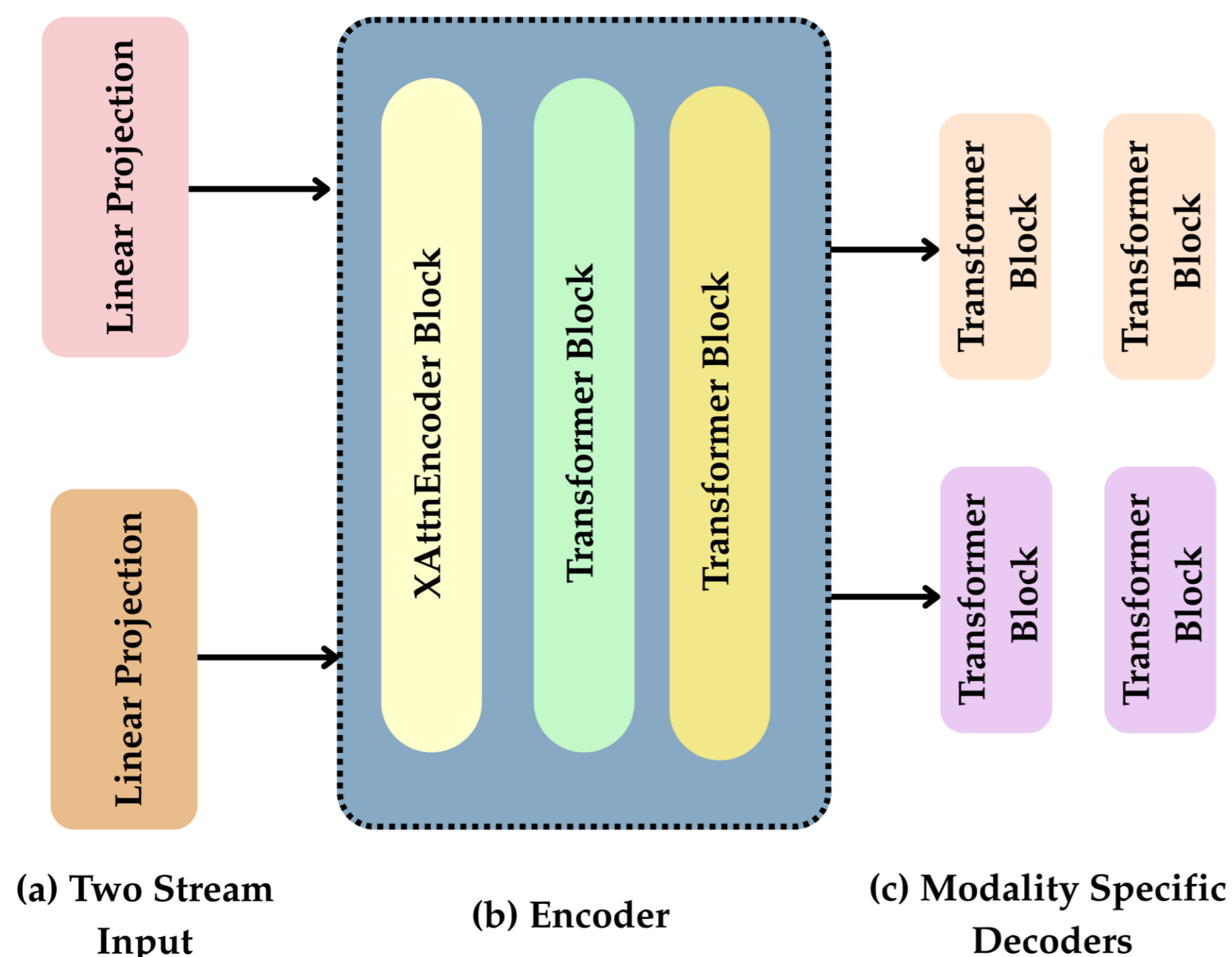


Fig. 1. Overall architecture of FusionX-Net

RESULTS & DISCUSSION

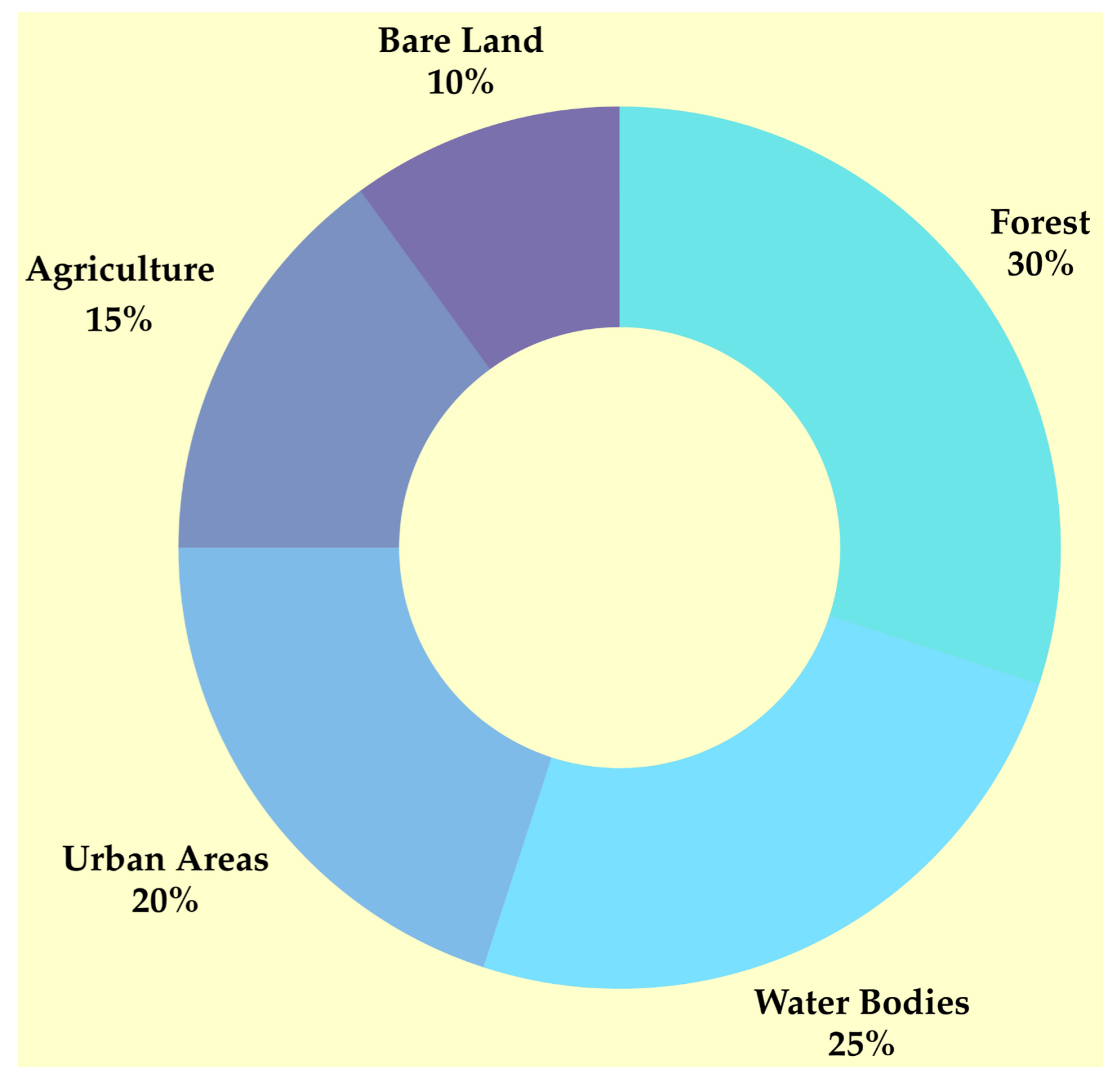


Fig. 2. Class Distribution for Each Dataset

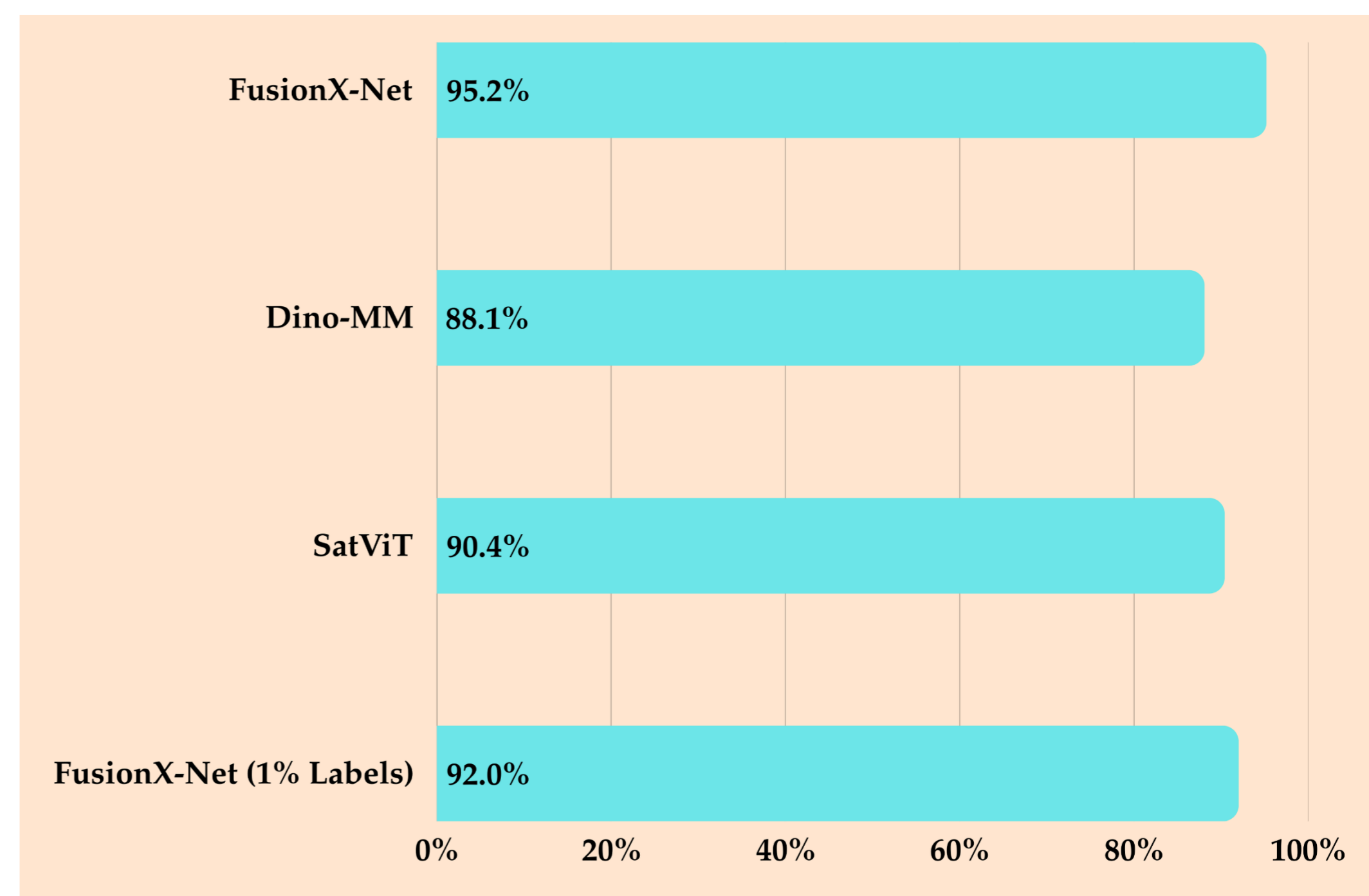


Fig. 3. Comparison with Existing Methods

CONCLUSION

- ❖ FusionX-Net fuses SAR and optical data efficiently using masked **autoencoders and cross-attention**.
- ❖ It **outperforms existing methods**, especially in low-label regimes, showing strong performance with limited labeled data.
- ❖ FusionX-Net provides **a scalable solution** for multi-modal data fusion in remote sensing without relying on complex data augmentation.