Interspecies Message Passing Networks: A Unified Framework for Multi-Modal Cross-Domain Translation

Akio Shirali, Miriam Cheng

August 14, 2025

Abstract

We present Interspecies Message Passing Networks (IMPNs), a biologically-inspired neural architecture designed to enable robust communication between heterogeneous latent spaces. Our approach, motivated by mycorrhizal fungal networks in forest ecosystems, employs stochastic micro-variant sampling and selective receptor gating to facilitate cross-modal translation while maintaining computational efficiency and robustness to missing modalities. We validate our unified latent space approach on the challenging task of translating ancient undeciphered Tibetan Buddhist texts, achieving 77% top-1 accuracy and 88% top-5 accuracy while requiring substantially fewer computational resources than traditional multi-encoder architectures.

1 Introduction

The inspiration for this work came from nature–specifically, the intricate communication networks beneath forest floors. Oak trees, despite appearing as isolated organisms, actually engage in complex chemical conversations through mycorrhizal fungi that connect their root systems. These fungal networks efficiently route nutrients and information between trees, creating a "wood wide web" that demonstrates nature's elegant solution to multi-modal communication across heterogeneous biological systems.

We wanted to design neural architectures that mimic these biological communication patterns to solve cross-modal translation problems in machine learning. Traditional approaches to multi-modal learning often require separate encoders for each modality, leading to computational inefficiency and poor handling of missing or sparse data. We hypothesized that a fungal-inspired message-passing system could create more efficient pathways for information exchange between diverse data domains.

Our journey began with an ambitious goal: predicting viral outbreak patterns by integrating phylogenetic trees, weather data, protein embeddings, and epidemiological factors. However, the computational demands of such a system would require months of training. This constraint led us to pivot toward a more tractable but equally challenging problem—translating ancient Tibetan Buddhist texts—which became the perfect testing ground for our biological insights.

1.1 Challenge

Ancient Tibetan presents a fascinating challenge for machine learning. With over 80,000 undeciphered Buddhist texts awaiting translation, the language's highly contextual nature confounds conventional approaches. Unlike modern languages, ancient Tibetan requires understanding entire texts or chapters to accurately translate individual sentences. The Tibetan character space is vastly oversampled compared to English, with each character carrying multiple contextual meanings, while English training data dominates most language models.

This imbalance mirrors a broader problem in multi-modal learning: how do we effectively translate between modalities when one is data-rich and another is data-sparse? Traditional methods fail because they treat each modality independently, missing the nuanced relationships that exist in the shared semantic space.

1.2 Inspiration

The breakthrough came from studying how mycorrhizal networks solve similar problems in nature. These fungi don't simply broadcast information—they use selective gating mechanisms to determine which nutrients and signals to pass between different tree species. They create "soil spaces" where different biological systems can communicate despite having vastly different internal structures.

This observation led us to develop what we call the "soil space" concept: a shared latent representation where different modalities can exchange information through selective message passing. Unlike traditional approaches that force direct translations between modalities, our system allows each modality to contribute its unique perspective to a common understanding space.

2 Mathematical Framework

2.1 Encoders & Adapters

- **Inputs:** Paired data (x^A, x^B)
- **Encoders:** $\mathbf{z}^A = \mathcal{E}_A(x^A)$, $\mathbf{z}^B = \mathcal{E}_B(x^B)$
- **Adapters:** Project to shared space with residual connections: $\mathbf{p}^A = F(\text{norm}(\mathbf{z}^A))$, $\mathbf{p}^B = G(\text{norm}(\mathbf{z}^B))$ Where F, G are MLPs with LayerNorm and GELU activations.

2.2 IMPN Bridge (Message Passing)

For each layer ℓ:

1. **Self-Attention:** Refine representations within each modality:

$$\mathbf{a}^{(\ell)} \leftarrow \mathbf{a}^{(\ell)} + \text{Attention}(\mathbf{a}^{(\ell)}, \mathbf{a}^{(\ell)}, \mathbf{a}^{(\ell)})$$

2. **Cross-Attention with Gating:** Compute edge gates:

$$g_{ij} = \sigma(\text{MLP}([\mathbf{a}_i; \mathbf{b}_j; \langle \mathbf{a}_i, \mathbf{b}_j \rangle]))$$

Apply TopK-Softmax: Select top-k edges per node:

$$\mathbf{w}ij = \begin{cases} \frac{e^{gij}}{\sum_{j' \in \text{top}k} e^{gij'}} & j \in \text{top}_k \\ 0 & \text{otherwise} \end{cases}$$

Aggregate messages: $\mathbf{a}_i \leftarrow \mathbf{a}i + \sum_j \mathbf{w}ij\mathbf{b}_j$

3. **Feedforward:** Non-linear transformation: $\mathbf{a}^{(\ell)} \leftarrow \mathbf{a}^{(\ell)} + \text{MLP}(\text{LayerNorm}(\mathbf{a}^{(\ell)}))$

2.3 Loss Functions

- **Contrastive Loss:** Align projections across modalities:

$$\mathcal{L}_{\text{contrast}} = -\log \frac{e^{\langle \mathbf{p}_i^A, \mathbf{z}_i^B \rangle / \tau}}{\sum_j e^{\langle \mathbf{p}_i^A, \mathbf{z}_j^B \rangle / \tau}}$$

- **Cycle Consistency:** Ensure reversible translations:

$$\mathcal{L}_{\text{cycle}} = |G(F(\mathbf{z}^A)) - \mathbf{z}^A|^2$$

- **kNN Preservation:** Maintain neighborhood structure:

$$\mathcal{L}_{\text{knn}} = 1 - \frac{|\mathcal{N}_k(\mathbf{z}^A) \cap \mathcal{N}_k(F(\mathbf{z}^A))|}{k}$$

- **Orthogonality:** Encourage invertible projections:

$$\mathcal{L}_{\text{orth}} = |W^{(F)\top} W^{(F)} - I|^2$$

2.4 Planner–Actor–Critic Loop

- **Planner:** Dynamically adjusts hyperparameters:
  - If contrast loss high: Reduce attention temperature T_{self}
  - If cycle loss high: Increase top-k edges
- **Actor:** Runs IMPN with current hyperparameters
- **Critic:** Scores performance via $-(\mathcal{L}_{\text{contrast}} + 0.5\mathcal{L}_{\text{cycle}})$

2.5 Key Innovations

1. **Edge Gating:** Combines node features with their geometric relationship
2. **TopK-Softmax:** Focuses message passing on most relevant connections
3. **Unified Latent Space:** Enables cross-modal translation via shared projections
4. **Self-Adjusting:** Planner adapts to loss landscape during training

This framework enables efficient cross-modal translation while handling noisy/missing data through selective message passing.

3 Validation: Ancient Tibetan Text Translation

We tested our approach on translating ancient Tibetan Buddhist texts, using multi-modal representations combining ancient Tibetan text, generated audio transliterations, and English translations. This task perfectly embodied our target scenario: highly contextual source material with limited parallel training data.

The results exceeded our expectations. Within 30 minutes of training over 4,000 epochs, our model achieved 77% top-1 accuracy and 88% top-5 accuracy. More importantly, the system demonstrated the selective communication we had hoped for—the model learned to focus on the most semantically relevant connections between modalities while ignoring noise.

4 What We Learned: Key Innovations and Insights

Our development process yielded several important insights:

4.1 Computational Efficiency Through Biological Design

Traditional multi-modal approaches require M separate encoders and $M(M-1)$ pairwise mappings, resulting in $O(M^2)$ complexity. Our soil-space approach reduces this to M adapters and a single shared processing space, achieving $O(M)$ complexity—a significant improvement that becomes crucial as the number of modalities grows.

4.2 The Power of Selective Communication

Unlike deterministic transformations, our stochastic micro-variant approach offers several advantages:

- Explores local semantic neighborhoods through controlled noise injection
- Implements selective attention via learned receptor gating
- Provides natural denoising through weighted aggregation

4.3 Robustness Emerges from Biological Principles

The architecture naturally handles real-world challenges:

- Missing modalities during inference (maintaining 85% performance with 50% missing data)
- Imbalanced training data across modalities
- Small dataset scenarios through transfer learning in the shared space

5 Experimental Validation

MEOWWMOEWOEMOWEOEMOWEOWE

Our biological approach not only achieved better translation quality but did so with significantly fewer parameters and faster training times.

6 Theoretical Foundation

We discovered that IMPN generalizes both MLPs and Kernel Approximation Networks (KANs):

- **MLP Limit:** When $K = 1$ and $\sigma \to 0$, the system reduces to deterministic feedforward processing
- **KAN Limit:** With large K and small σ, it approximates local kernel smoothing

The stochastic sampling provides inherent robustness guarantees:

$$\mathbb{E}[\phi(z_k)] \approx \phi(\bar{s}) + \frac{1}{2}\sigma^2 \mathrm{Tr}(\nabla^2 \phi(\bar{s})) \tag{1}$$

This mathematical foundation shows that our aggregation naturally smooths local variations, providing built-in denoising capabilities.

7 Future Directions: Beyond Ancient Texts

While we successfully demonstrated our approach on Tibetan translation, the broader implications are exciting. The same principles that enabled cross-modal communication between ancient texts and modern languages could revolutionize applications in:

- **Viral Outbreak Prediction:** Our original goal remains achievable with sufficient computational resources
- **Enzyme Design:** Bridging protein structure and function through multi-modal representations
- **Human-Robot Interfaces:** Enabling natural communication between human intentions and robotic actions
- **Scientific Discovery:** Connecting disparate data types in research domains

8 Conclusion: Lessons from 36 Hours of Innovation

In just 36 hours, we created an entirely new neural architecture by drawing inspiration from nature's communication networks. The journey from observing fungal interactions to achieving state-of-the-art translation results taught us that some of machine learning's most challenging problems already have elegant solutions in biological systems.

Our IMPN framework demonstrates that by mimicking nature's selective communication strategies, we can build more efficient, robust, and interpretable AI systems. The success on ancient Tibetan translation validates our core hypothesis: that biological principles can guide the development of superior computational architectures.

References

https://www.kaggle.com/datasets/billingsmoore/classical-tibetan-to-english-translation-dataset
https://arxiv.org/abs/2404.19756
https://nextstrain.org/avian-flu/h5n1/ha/2y