From Graphs to Images: Non-Parametric PPI Context Integration for Vision-Based Protein Function Prediction

Yeonatan Mauhnoom[1][a], Gabriel Bianchin de Oliveira[1,2][b], Helio Pedrini[1][c], and Zanoni Dias[1][d]

[1]*Institute of Computing, University of Campinas, Campinas, Brazil, 13083-052*

[2]*Knight Foundation School of Computing and Information Sciences, Florida International University, Miami, USA, 33172*

yeonatan.mauhnoom@students.ic.unicamp.br, {gabriel.oliveira, helio, zanoni}@ic.unicamp.br

Keywords: Protein Function Prediction, Protein-Protein Interactions, Graph Neural Networks, Protein Language Models, Visual Representation Learning, Gene Ontology Annotation, Deep Learning, Bioinformatics

Abstract: Protein function annotation is critical for biomedical discovery, yet most proteins remain unannotated. Recent protein language models provide universal sequence embeddings; however, leveraging protein–protein interaction (PPI) networks alongside these embeddings remains challenging. PPI networks are inherently sparse and noisy, especially for understudied organisms, while existing approaches typically rely on parametric graph neural networks, introducing computational complexity and interpretability challenges. We propose a fixed, non-parametric approach that contextualizes sequence embeddings via multi-hop message passing over PPI graphs, then encodes enriched representations as RGB images for vision models. This strategy avoids the computational overhead of complex GNN architectures and yields interpretable predictions grounded in explicit image features. We evaluate our method against three baselines using CAFA5 data across three Gene Ontology domains. Our approach achieves wF$_{\text{max}}$ of 53.21% on BP, 62.98% on CC, and 68.97% on MF, exceeding DeepNF by more than 25 percentage points. The source code is publicly available at `https://github.com/ymauhn/PPI-context-integration-for-Protein-Function-Prediction`.

1 Introduction

Proteins are macromolecules composed of linear chains of amino acids that fold into specific three-dimensional structures, acquiring conformations that determine their biological functions (Satyanarayana, 2013; Chua et al., 2024). They are involved in almost every cellular process in living organisms, playing essential roles including catalysis of biochemical reactions, gene expression regulation, cell signaling, molecular transport, immune defense, and structural support (Anfinsen, 1973).

Functional protein annotation is organized in ontology, with the most utilized being the Gene Ontology (GO), subdivided into three main domains: Biological Process (BP), which relates to the cellular processes a protein participates in; Molecular Function (MF), describing elementary biochemical activities performed by the protein; and Cellular Component (CC), referring to the subcellular location where the protein acts or resides (Ashburner et al., 2000).

Understanding protein roles is crucial for advancing knowledge of biological systems and practical applications such as drug discovery and targeted therapies (Zhong et al., 2024). Although sequence databases such as UniProt and TrEMBL are growing rapidly (The UniProt Consortium, 2024), most known proteins still lack reliable functional annotations, mainly due to the cost and complexity of experimental validation.

To address the mismatch between the pace of sequence discovery and the slower, costly process of experimental annotation, automated function prediction methods based on Machine Learning (ML) and Deep Learning (DL) have emerged. Their importance has been reinforced by the Critical Assessment of Functional Annotation (CAFA), the main community challenge for protein function prediction, where recent editions show neural models surpassing traditional approaches across GO domains (Zhou et al., 2019; Radivojac et al., 2013).

Deep learning annotation methods have evolved along several lines (Boadu et al., 2025). Sequence-

[a] https://orcid.org/0009-0009-0362-2861
[b] https://orcid.org/0000-0002-1238-4860
[c] https://orcid.org/0000-0003-0125-630X
[d] https://orcid.org/0000-0003-3333-6822

based approaches, such as protein language models (Rives et al., 2021; Elnaggar et al., 2021), now surpass traditional alignment (e.g. BLAST and DIAMOND (Altschul et al., 1997; Buchfink et al., 2015)) tools by capturing biochemical, structural, and functional properties in vector embeddings.

Another direction relies on 3D structural models, catalyzed by advances such as AlphaFold, which enabled large-scale, high-accuracy structure prediction (Jumper et al., 2021). This structural information is, in turn, used to infer function, for instance, by identifying structural similarity to proteins of known function or by analyzing conserved active sites.

Finally, methods based on Protein-Protein Interaction (PPI) networks have gained momentum by leveraging the 'guilt-by-association' (or function transfer) principle, which assumes that proteins with functional or physical interactions tend to share similar biological roles (Piovesan et al., 2015).

Using PPI networks alone presents significant limitations: interaction databases such as STRING (Szklarczyk et al., 2021) integrate multiple evidence sources with varying reliability, introducing noise (You et al., 2021). Furthermore, many proteins, especially from under-studied organisms, have few known interactions, resulting in critical sparsity that weakens network-based methods (Boadu et al., 2025). Despite these challenges, recent evidence indicates that information from PPI graphs complements sequence-derived features, especially for biological processes and complexes not evident from sequence alone (Barot et al., 2021; Gillis and Pavlidis, 2011).

PPI-based function prediction spans network-only to multimodal approaches. Classical methods such as DeepNF (Gligorijevic et al., 2018) and NetQuilt (Barot et al., 2021) propagate signals on graphs to obtain network embeddings. In DeepNF, multiple STRING evidence networks are first diffused via Random Walk with Restart (RWR); the resulting co-occurrence statistics are transformed into a Positive Pointwise Mutual Information (PPMI) matric, which are fed to deep autoencoders to learn low-dimensional protein embeddings. These embeddings are then used by an SVM to predict Gene Ontology terms. However, these purely PPI-based methods remain fundamentally limited by the aforementioned sparsity and noise.

Multimodal models, such as DeepGO (You et al., 2018), revealed gains by complementing graph context with sequence embeddings, a trend accelerated by modern protein language models, that provide robust node features for graph-based approaches (Fan et al., 2020; Xu et al., 2024).

Moreover, advanced approaches, such as ProtHGT (Ulusoy and Dogan, 2025) and SEGT-GO (Wang et al., 2025), leverage Transformer-based graph models to exploit multi-relational biological context. ProtHGT builds a large heterogeneous knowledge graph integrating diverse entity types (e.g., proteins, GO terms, domains, diseases, phenotypes) and relations (protein–GO links, PPIs, orthology, disease–phenotype, protein–drug, etc.), and applies Heterogeneous Graph Transformer (HGT) layers with type-specific projections and edge-type attention; proteins are initialized with protein-language-model embeddings, while other nodes use fit-for-purpose embeddings (e.g., anc2vec for GO, dom2vec for domains). SEGT-GO focuses on PPI graphs and sequence-derived InterPro features, first performing non-parametric multi-hop serialization via repeated multiplication of the normalized adjacency with feature matrices to form hop-wise "tokens", then feeding these tokens to a Graph Transformer encoder with a multi-hop attention aggregator; a SHAP-based filter is used to reduce feature noise before classification.

While powerful, such methods can face (1) substantial computational overhead from deep attention-based architectures and (2) amplified feature noise when heterogeneous evidence (e.g., multi-channel STRING) is sparse or unreliable, which may degrade effectiveness in some settings (Chen et al., 2024).

In contrast, other state-of-the-art methods, such as SUPERMAGO (Oliveira et al., 2025), focus solely on sequence signals, exploiting rich embeddings from large-scale language models. While such approaches deliver strong results (particularly for MF), they inherently ignore the critical functional context encoded in protein interactions, struggling to capture systemic roles, such as BP or CC, as these functions are defined by their relationships and interactions with other proteins. Systematic assessment of this synergy requires methods that can handle PPI sparsity and heterogeneity, leverage multi-hop neighborhood information, and do so without the excessive computational burden of deep GNNs.

In this work, we address these challenges by combining topological context from PPI networks with robust ProtT5 sequence embeddings, applying fixed, non-parametric message passing prior to visual encoding. Rather than feeding vectors directly into classifiers or training complex GNNs, we transform contextualized embeddings into RGB images for deep vision backbones.

To evaluate our approach, we compare three baselines: (i) a visual baseline using grayscale images from score vectors and a convolutional backbone; (ii) DeepNF, a state-of-the-art PPI-based method; (iii) a simple non-parametric PPI-based baseline using score-weighted label transfer. Our proposed image-based approach outperforms both baselines across all domains, reaching 53.21%, 62.98% and 68.97% of $\mathrm{wF}_{\max}$ on BP, CC and MF, respectively.

Three main contributions of this research are: (1) the proposal of a non-parametric, multi-hop message passing approach to contextualize sequence embeddings with PPI network topology via controlled hyperparameters, eliminating the computational overhead and interpretability barriers of complex Graph Neural Network architectures; (2) the introduction of a novel visual encoding strategy that transforms PPI-contextualized embeddings into RGB images, enabling the application of powerful pretrained vision backbones to protein function prediction; and (3) results presenting robust method performance across heterogeneous and sparse PPI networks, effectively handling under-studied organisms where traditional network-based methods fail due to incomplete interaction coverage.

This paper is organized as follows. Section 2 describes the proposed method, datasets, evaluation metrics and comparison approaches. Section 3 presents results for BP, MF, and CC. Section 4 synthesizes key findings and outlines future directions.

2 Materials and Methods

In this section, we describe the CAFA5-based dataset and STRING network preprocessing, outline ProtT5 embedding extraction and our image representation approach, specify the loss and training setup, define evaluation metrics, and summarize the comparison baselines.

2.1 Proposed Method

Our approach transforms protein-protein interaction context into visual representations suitable for deep learning-based classification of GO functional annotations. The overall pipeline, presented in Figure 1, consists of four main stages: (i) collection and preprocessing of PPI networks and protein sequences; (ii) generation of image representations encoding PPI-contextualized sequence embeddings via message passing; (iii) fine-tuning of pretrained vision backbones for multi-label GO term prediction; and (iv) hierarchical aggregation via the True Path Rule (TPR) to ensure consistency with the GO directed acyclic graph structure.

A core aspect of our approach is that embedding enrichment with graph context is carried out through a non-parametric preprocessing step that mimics message passing in Graph Neural Networks, but without requiring supervised learning at this stage.

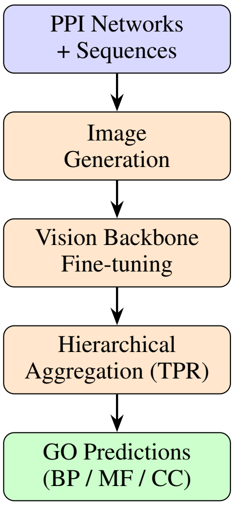


Figure 1: High-level pipeline for protein function prediction via visual representations. PPI networks and protein sequences are transformed into RGB images through message passing, which are then used to fine-tune a vision backbone. Predictions are hierarchically aggregated using the True Path Rule to ensure GO DAG consistency.

This formulation offers interpretable control via explicit hyperparameters: the number of hops defines the radius of contextual diffusion in the graph; the interaction-score threshold filters noise by prioritizing high-confidence edges; and the coefficient α tunes the balance between the original protein embedding and the propagated neighborhood vectors, allowing a smooth spectrum from self-dominated representations ($\alpha \approx 1$) to context-dominated ones ($\alpha \approx 0$). With that, our model provides an interpretable and scalable alternative to complex deep GNN architectures while maintaining robustness to PPI sparsity and noise.

We extract sequence-based representations using ProtT5-XL-U50 (Elnaggar et al., 2021), a transformer-based protein language model pretrained on over 2.1 billion protein sequences from UniRef50 (Suzek et al., 2007). For each protein, we compute embeddings from the last three encoder layers (layers 22, 23, and 24), each with dimension 1024.

To handle proteins longer than the model's input limit, sequences exceeding 1022 residues are split into non-overlapping chunks. Each chunk is independently processed through ProtT5, producing per-residue hidden states. For each layer, we apply attention-masked mean pooling over residues to obtain a fixed-size 1024-dimensional vector per chunk. When a protein is split into multiple chunks, the resulting vectors are averaged across chunks to yield a single 1024-dimensional representation per layer. This chunking strategy ensures all proteins, regardless of length, receive comparable treatment without truncation.

The three layer-wise embeddings (layers 22, 23, 24) are stored separately and later combined during image generation. All embeddings are precomputed offline using mixed-precision inference (bfloat16) to reduce memory overhead and accelerate extraction across the full dataset.

Our proposed method enriches ProtT5 sequence embeddings with multi-hop neighborhood context via a fixed, non-parametric message-passing procedure. This approach constructs a normalized graph and iteratively propagates information across multiple hops, enabling the model to capture multi-step functional associations encoded in PPI topology.

Let $\mathcal{V}$ denote the set of proteins ($|\mathcal{V}| = n$). We construct a weighted adjacency matrix $W \in \mathbb{R}^{n\times n}$ from STRING combined scores by: (1) retaining protein pairs (u, v) with combined score $\max(s_{uv}, s_{vu})$ is positive; (2) excluding edges where $u = v$ (no self-loops); (3) deduplicating entries so each unordered pair $\{u, v\}$ appears at most once with the maximum score; (4) ensuring bidirectional edges by setting $W_{uv} = W_{vu} = \max(s_{uv}, s_{vu})$; and (5) computing column sums $(D_c)_{jj} = \sum_i W_{ij}$ to normalize columns to sum 1, defining the column-stochastic transition matrix $P = W D_c^{-1}$. This normalization ensures that aggregated embeddings at each hop are weighted by the relative importance of each neighbor within its local neighborhood, preventing dominance by high-degree hubs.

We concatenate the three per-layer ProtT5 embeddings (layers 22, 23, 24) to form a unified 3072-dimensional base representation $H^{(0)} \in \mathbb{R}^{n\times 3072}$ for all n proteins. We then apply H iterations of fixed message passing to this concatenated representation:

$$H^{(h)} = \alpha H^{(h-1)} + (1-\alpha) P H^{(h-1)}, \quad h = 1, \ldots, H,$$

where P is the column-stochastic transition matrix, $\alpha \in [0, 1]$ balances self-information and neighborhood aggregation, and H determines the propagation radius. Unlike processing each layer independently, this formulation propagates joint information across all three ProtT5 layers simultaneously, allowing interactions between layer-specific features during the diffusion process.

Hyperparameters H and α are tuned per ontology on the validation set. For each protein, we extract the corresponding row from the final contextualized embedding matrix $H^{(H)} \in \mathbb{R}^{n\times 3072}$, yielding a 3072-dimensional vector. This vector is clipped to the range $[-50, 50]$ to prevent numerical instability, then passed through a sigmoid activation to map values to $[0, 1]$. The resulting normalized vector is reshaped directly into a $3 \times 32 \times 32$ tensor, where the first 1024 elements form the red channel, the next 1024 form the green channel, and the final 1024 form the blue channel. For storage, pixel values are scaled to $[0, 255]$ and cast to 8-bit integers.

We employ pretrained convolutional neural network architectures as vision backbones to classify the generated protein images. Following the approach demonstrated in (Oliveira et al., 2025), we primarily use ResNet-50 (He et al., 2016), a residual network with 50 layers pretrained on ImageNet. While ResNet-50 serves as our primary backbone, we also evaluate ConvNeXt-Tiny (Liu et al., 2022) in ablation studies to assess robustness to architectural choices.

The original 1000-class classification head is replaced with a domain-specific multi-label head consisting of a single fully connected layer with C outputs, where C corresponds to the number of retained GO terms in the target ontology (BP, CC, or MF). Each output is passed through a sigmoid activation to produce independent probabilities for each GO term.

We evaluated multiple fine-tuning strategies during model development: freezing all backbone parameters and training only the classification head, partial fine-tuning (unfreezing only the final blocks), and full end-to-end fine-tuning. Based on validation performance, we adopted full backbone fine-tuning for all reported experiments, allowing the model to adapt pretrained features to the protein domain. Notably, while the backbones are pretrained on ImageNet, we found that applying ImageNet normalization (channel-wise mean/std standardization) to our protein images degraded performance; consequently, images are fed to the network without additional normalization beyond the sigmoid scaling applied during image construction.

All models are trained using the Adam optimizer with an initial learning rate of 1×10^{-4}. We use a batch size of 64. The training was conducted with early stopping with patience of 20 epochs. Random seeds are fixed for reproducibility.

We adopt Protein Loss, an information-weighted composite loss designed to address class imbalance and capture dual perspectives (protein-centric and GO-centric) in multi-label prediction (Liu et al., 2024). Protein Loss is defined as:

$$\mathcal{L}_{\text{Protein}} = \mathcal{L}_{\text{ZLPR}} \times \mathcal{L}_{\text{F1}}^{\text{GO}} \times \mathcal{L}_{\text{F1}}^{\text{protein}},$$

where each component addresses a distinct challenge in multi-label GO prediction. The first component, $\mathcal{L}_{\text{ZLPR}}$ (Su et al., 2022), is a ranking-based categorical cross-entropy loss that encourages the model to assign higher scores to positive labels than negative labels via pairwise margin constraints. This formulation naturally handles class imbalance by focusing on relative ranking rather than absolute probabilities. The second and third components are IC-weighted F1 losses that operate along complementary dimensions of the prediction matrix. Let $w_j = \mathrm{IC}(j)$ denote the Information Content (IC) of term j, defined as:

$$\mathrm{IC}(j) = -\log \frac{\mathrm{freq}(j)}{N},$$

where $\mathrm{freq}(j)$ is the number of proteins annotated with term j (including ancestors via the True Path Rule) and N is the total number of proteins. The GO-centric F1 loss $\mathcal{L}_{\mathrm{F1}}^{\mathrm{GO}}$ aggregates IC-weighted precision and recall across proteins for each term, then averages over terms. Conversely, the protein-centric F1 loss $\mathcal{L}_{\mathrm{F1}}^{\mathrm{protein}}$ aggregates IC-weighted metrics across terms for each protein, then averages over proteins. This dual formulation ensures the model balances performance both per-term (avoiding neglect of rare labels) and per-protein (ensuring consistent multi-label predictions for individual instances).

GO annotation is inherently multi-label and hierarchical, structured as a directed acyclic graph where terms are linked by is-a and part-of relationships. The True Path Rule (Valentini, 2010) enforces the biological constraint that if a protein is annotated with a specific term, it must also be annotated with all ancestor terms in the DAG. To maintain hierarchical consistency, predictions for each protein are closed upward along the GO DAG, so if a child term is predicted above threshold, all its ancestors are also set to positive.

2.2 Datasets

We use protein–protein interaction data from STRING (Szklarczyk et al., 2021), which integrates evidence from seven channels: genomic neighborhood, gene fusion, phylogenetic co-occurrence, co-expression, curated databases, experimental data, and text mining. Each protein pair is assigned a combined confidence score (global score) ranging from 0 to 1. We construct the edge set from all pairs with strictly positive scores.

We use the protein datasets from MAGO (Oliveira et al., 2024), which retains only the most frequent GO terms in each ontology domain (BP, CC, MF). Table 1 summarizes the dataset sizes and term counts per domain. Specifically, the table reports the number of proteins in the training, validation, and test splits, as well as the number of distinct GO terms (labels) for each ontology.

Table 1: Dataset statistics per GO ontology. Columns report the number of distinct proteins in train, validation, and test splits, and the number of GO terms retained after filtering for each ontology.

Ontology	Train	Validation	Test	Terms
BP	73,768	9,221	9,221	500
CC	74,328	9,292	9,292	498
MF	62,909	7,864	7,864	499

2.3 Evaluation Metrics

We evaluate all methods using the official CAFA metrics (Zhou et al., 2019), which are specifically designed for multi-label, hierarchical protein function prediction. The $\mathrm{wF}_{\max}$ metric evaluates predictions across different thresholds $\tau \in (0,1]$, finding the threshold where the harmonic mean of weighted precision and recall reaches its maximum:

$$\mathrm{wF}_{\max} = \max_{\tau} \left\{ \frac{2 \cdot pr(\tau) \cdot rc(\tau)}{pr(\tau) + rc(\tau)} \right\},$$

where:

$$pr(\tau) = \frac{1}{m(\tau)} \sum_{i=1}^{m(\tau)} \frac{|P_i(\tau) \cap T_i| \cdot \mathrm{IC}}{|P_i(\tau)| \cdot \mathrm{IC}},$$

$$rc(\tau) = \frac{1}{n} \sum_{i=1}^{n} \frac{|P_i(\tau) \cap T_i| \cdot \mathrm{IC}}{|T_i| \cdot \mathrm{IC}}.$$

Here, $P_i(\tau)$ is the set of predicted terms with score $\geq \tau$ for protein i, T_i is the set of true terms for protein i, $m(\tau)$ is the number of proteins with at least one prediction at threshold τ, and n is the total number of proteins. This metric emphasizes rare, specific terms by weighting each term's contribution according to its information content.

2.4 Comparison Approaches

To evaluate the effectiveness of our image-based approach, we compare against three baselines representing different paradigms for leveraging PPI information in protein function prediction. The nonparametric naive baseline directly aggregates neighbor annotations via PPI score-weighted label transfer, without any trainable parameters. For each protein v, let $s_{uv} \in [0,1]$ denote its combined PPI score with training protein u. We normalize these scores to form weights:

$$w_u = \frac{s_{uv}}{\sum_{u' \in \mathcal{N}(v)} s_{u'v}},$$

where $\mathcal{N}(v)$ is the set of neighbors with nonzero interaction scores. The predicted probability for GO term j is then:

$$\hat{y}_j(v) = \sum_{u \in \mathcal{N}(v)} w_u \cdot y_j(u),$$

where $y_j(u) \in \{0,1\}$ indicates whether term j is annotated to protein u.

DeepNF (Gligorijevic et al., 2018) is a deep autoencoder framework originally designed to integrate heterogeneous biological networks (representing different evidence types) into compact low-dimensional protein embeddings. The method preprocesses each network via Random Walk with Restart (RWR) followed by Positive Pointwise Mutual Information (PPMI) transformation, producing high-dimensional representations that capture higher-order network proximities.

In our implementation, we adapt DeepNF in several ways to match our experimental setup. First, since our data uses only STRING's global (combined) score rather than the seven individual evidence channels, we employ the single-network autoencoder variant of DeepNF. Second, given that our dataset contains approximately 500 GO terms per ontology, significantly more labels than the original DeepNF study, we replace the SVM classifier with a 3-layer MLP (hidden dimensions: 1024, 512, 256; dropout 0.3) to handle the increased output dimensionality and improve scalability. Third, to enable fair comparison with our method, we train the MLP classifier using Protein Loss (the same loss function as our method). All other hyperparameters (RWR restart probability, PPMI construction, autoencoder architecture) follow the DeepNF paper.

The visual PPI-only baseline generates grayscale images directly from STRING combined confidence scores without incorporating sequence information or multi-hop aggregation. For each protein, we construct a raw interaction vector containing the combined (global) score for all training proteins.

Training proteins include a self-interaction score set to 1.0, while validation and test proteins include scores for all training proteins plus themselves, yielding vectors of dimension N (training) or $N+1$ (validation/test), where N is the training set size. We apply deduplication to ensure each protein pair appears at most once. To produce a fixed-size representation, we apply max-pooling to reduce the interaction vector to dimension 224, yielding a compressed profile $\mathbf{v} \in \mathbb{R}^{224}$. We then construct a 224×224 dissimilarity matrix via outer absolute differences:

$$M_{ij} = |v_i - v_j|,$$

resulting in a symmetric matrix with zero diagonal. This matrix is normalized to $[0,1]$ via min-max scaling and treated as a single-channel (grayscale) image.

This representation encodes local variation patterns in the protein's interaction profile but captures only direct (first-order) neighborhood information.

All three baselines, along with our proposed method in its best hyperparameter configuration (selected via validation set performance), are evaluated on the held-out test split using the metrics described in Section 2.3.

3 Results and Discussion

In this section, we present a systematic ablation study quantifying the contribution of each component in our pipeline. The experimental sequence progresses as follows: (1) multi-hop ablation across domains; (2) α-weight optimization at domain-specific hop counts; (3) vision backbone comparison; and (4) final benchmark against all baselines.

3.1 Multi-Hop Analysis

In this first analysis, we vary the number of message-passing hops H from 0 to 10 for each ontology, using ResNet50 and fixed $\alpha = 0.5$.

Table 2 presents the results achieved by each hop configuration. Multi-hop aggregation shows domain-specific optima. BP and CC rise up to $H = 5$ (49.68%) and $H = 6$ (59.87%), respectively, yielding +5.89pp and +4.59pp over $H = 0$, then decline, consistent with over-smoothing, where deeper message passing makes node embeddings less distinguishable. In contrast, MF peaks at $H = 0$ (63.71%) and degrades steadily to 56.83% at $H = 10$, indicating sequence-driven signals that PPI aggregation may dilute via over-smoothing and noisy neighborhood propagation. We therefore adopt $H = 5$ (BP), $H = 6$ (CC), and $H = 0$ (MF) henceforth.

Table 2: Multi-hop analysis using $\mathrm{wF_{max}}$ on the validation set.

Hops (H)	BP	CC	MF
0	43.79	55.28	**63.71**
1	41.73	52.59	57.55
2	44.33	54.88	57.63
3	46.52	57.30	57.94
4	47.72	57.96	59.37
5	**49.68**	59.09	59.87
6	49.15	**59.87**	60.52
7	47.20	58.41	60.12
8	45.88	57.13	59.75
9	44.51	55.69	58.94
10	43.12	54.20	56.83

3.2 Alpha-Weight Analysis

Using a ResNet-50 backbone and fixing the per-domain hop depth at the best values found (BP: H=5; CC: H=6; MF: H=0), we tune the interpolation weight α, which trades off sequence information (higher α) against PPI-based neighbor aggregation (lower α). For MF, α has no effect because H=0 disables message passing.

As shown in Table 3, the optimal α parameters for BP and CC are 0.5 and 0.4, respectively. Lower α values emphasize neighbor aggregation but carry the risk of over-smoothing, while higher α values preserve sequence information but may underutilize PPI context. The approximately balanced contributions of sequence and network context suggest that protein function prediction benefits from leveraging both intrinsic and relational signals.

Table 3: α-weight ablation at domain-specific optimal hop depths using wF$_{\text{max}}$ on the validation set.

α	BP	CC	MF
0.1	40.85	52.15	—
0.2	40.59	53.13	—
0.3	48.30	56.31	—
0.4	49.63	**60.43**	—
0.5	**49.68**	60.00	—
0.6	47.97	60.38	—
0.7	48.27	59.14	—
0.8	43.77	58.91	—
0.9	40.94	55.68	—

3.3 Vision Backbone Comparison

Next, using the optimal hop depth and α per domain from the previous analyses, we compare ResNet50 with ConvNeXt Tiny.

The results are shown in Table 4. ConvNeXt Tiny outperforms ResNet50 across all ontologies. This hierarchical multi-scale architecture effectively captures spatial patterns in graph-derived visual representations. All subsequent experiments employ ConvNeXt Tiny.

Table 4: Vision backbone comparison at optimal hops and α per domain using wF$_{\text{max}}$ on the validation set.

Backbone	BP	CC	MF
ResNet50	49.68	60.43	63.71
ConvNeXt Tiny	**52.60**	**62.82**	**68.89**

3.4 Test Set Evaluation

As presented in Table 5, we compared our final model with the comparison approaches. Our method achieves test set wF$_{\text{max}}$ of 53.21% (BP), 62.98% (CC), and 68.97% (MF), substantially outperforming all baselines. Compared to the naive baseline, we gain 12.99pp (BP), 14.86pp (CC), and 24.83pp (MF). Relative to DeepNF, improvements are more substantial: 23.94pp (BP), 26.20pp (CC), and 25.09pp (MF). On average across GO domains, our method reaches 61.72% wF$_{\text{max}}$, exceeding the naive baseline by 17.56pp and DeepNF by 25.08pp. The substantial MF gain despite zero multi-hop contribution demonstrates ConvNeXt Tiny's effectiveness in extracting hierarchical features from pure ProtT5 embeddings without PPI propagation.

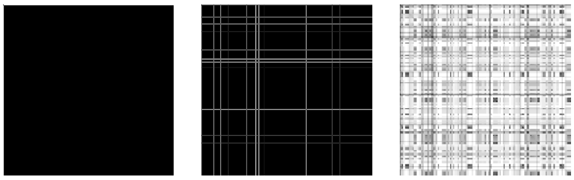
Table 5: Results using wF$_{\text{max}}$ on the test set.

Method	BP	CC	MF	Avg.
Naive baseline	40.22	48.12	44.14	44.16
DeepNF	29.27	36.78	43.88	36.64
PPI-only visual	28.44	39.81	40.65	36.30
Ours	**53.21**	**62.98**	**68.97**	**61.72**

3.5 Visual Analysis

Figure 2 demonstrates why PPI-only methods fail: isolated proteins (Q07623, with 0 neighbors) produce uninformative black images, while moderately connected proteins (Q99966, with 11 neighbors) and hub proteins (Q02248, with 351 neighbors) show increasing texture.

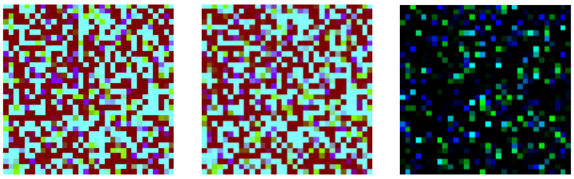
Our approach provides meaningful embeddings regardless of graph connectivity, since sequence information is globally informative. This explains superior performance on datasets with sparse PPI coverage, such as proteins from less-studied organisms.



(a) Q07623 (b) Q99966 (c) Q02248

Figure 2: PPI sparsity in visual encoding: isolated proteins produce uninformative black images (a), moderately connected proteins show sparse texture (b), hub proteins exhibit rich complexity (c).

Figure 3 shows that message passing enriches representations by gradually modulating mid-frequency patterns without disrupting sequence structure.



(a) $H = 0$ (b) $H = 1$ (c) difference

Figure 3: Message passing enrichment: (a) $H = 0$ pure sequence embedding, (b) $H = 1$ after one-hop PPI aggregation ($\alpha = 0.5$), (c) difference map showing spatial components modulated by neighborhood context.

3.6 Computational Considerations

Here, we summarize the time, memory, and storage costs of the non-parametric contextualization and image encoding steps. The graph contextualization stage is fixed and non-parametric (no trainable parameters and no backpropagation through edges). Let n be the number of proteins (nodes), $|E|$ the number of nonzero interactions (edges), d the embedding dimensionality (here $d = 3{,}072$ after concatenating the last three ProtT5 layers), and $P \in \mathbb{R}^{n\times n}$ the sparse column-stochastic transition matrix derived from STRING. For a chosen hop depth H (number of propagation steps), each hop updates the representation by a sparse matrix-dense matrix multiplication (e.g., $PH^{(h-1)}$), so the overall cost scales as $O(H \cdot |E| \cdot d)$ time in practice and $O(|E| + n \cdot d)$ memory to store the graph and embeddings.

The resulting $3 \times 32 \times 32$ images can be stored compactly as integers (using 8 bits), requiring 3,072 bytes (3 KB) per protein, whereas storing the same image as floating-point values (using 32 bits) would require 12,288 bytes (12 KB). Depending on storage constraints, images can also be generated on-the-fly from cached embeddings, reducing memory usage during training and inference.

4 Conclusion and Future Work

Protein function prediction is critical for computational biology but faces significant challenges in integrating protein–protein interaction (PPI) network data. A central gap lies in managing heterogeneous and imbalanced datasets, such as those in CAFA challenges, which aggregate proteins from diverse organisms with inherently sparse, incomplete, and noisy PPI coverage. Classical graph diffusion methods such as DeepNF, which perform adequately on dense networks from single organisms (e.g., human or yeast), prove overly sensitive to sparsity, failing to generalize across such heterogeneous settings. Conversely, state-of-the-art parametric Graph Neural Networks (GNNs), including GAT and GraphSAGE, though promising, often incur substantial computational overhead and interpretability challenges, limiting their scalability for large PPI networks.

We propose a robust and computationally lightweight non-parametric graph preprocessing stage that translates PPI context into visual representations via fixed message passing. This strategy avoids the architectural and hyperparameter-tuning complexity of trainable GNNs, while keeping neighborhood contextualization explicitly controlled and auditable through the number of hops (H) and the balance factor (α). Moreover, it empirically demonstrates improved robustness to data sparsity compared to network-only baselines.

Our results reveal gains exceeding 25 percentage points in wF$_{\text{max}}$ over DeepNF, highlighting the latter's limitations when faced with heterogeneous datasets. Furthermore, coupling these representations with modern CNN backbones (ConvNeXt Tiny) proves effective, capturing salient spatial patterns and exposing fundamental domain asymmetry: Molecular Function (MF) prediction is governed primarily by sequence information (0 hops), whereas Biological Process (BP) and Cellular Component (CC) prediction critically depend on network context (5-6 hops).

Limitations inherent in using fixed parameters frame clear directions for future research. The natural progression involves exploring learnable aggregation mechanisms. We plan to replace the fixed α with a learnable attention module allowing dynamic balancing of sequence and neighborhood information on a per-protein or evidence-type basis. Additionally, a key strength of our visual encoding lies in its interpretability potential; we intend to apply post-hoc visualization techniques such as Grad-CAM to elucidate which visual patterns and neighborhood contexts are most salient for specific predictions.

Finally, we plan to compare the visual prediction head against direct vector-based heads (e.g., an MLP or Transformer over the 3,072-dimensional embeddings). We also intend to conduct a rigorous runtime and resource benchmark against deep GNN and Graph Transformer alternatives. Channel-aware propagation (e.g., separate diffusion per STRING evidence type) and organism/group-specific analyses may further clarify when combined-score diffusion is most beneficial.

ACKNOWLEDGEMENTS

This work was supported by the São Paulo Research Foundation (grant number 2017/12646-3), the National Council for Scientific and Technological Development (grant numbers 302530/2022-3 and 304836/2022-2), and the Coordination for the Improvement of Higher Education Personnel.

REFERENCES

Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: A New Generation of Protein Database Search Programs. *Nucleic Acids Research*, 25(17):3389–3402.

Anfinsen, C. B. (1973). Principles That Govern the Folding of Protein Chains. *Science*, 181(4096):223–230.

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., Davis, A. P., Dolinski, K., Dwight, S. S., Eppig, J. T., Harris, M. A., Hill, D. P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J. C., Richardson, J. E., Ringwald, M., Rubin, G. M., and Sherlock, G. (2000). Gene Ontology: Tool for the Unification of Biology. *Nature Genetics*, 25(1):25–29.

Barot, M., Gligorijevic, V., Cho, K., and Bonneau, R. (2021). NetQuilt: deep multispecies network-based protein function prediction using homology-informed network similarity. *Bioinformatics*, 37(16):2414–2422.

Boadu, F., Cao, Y., and Cheng, J. (2025). Deep learning methods for protein function prediction. *Proteomics*, 25(1-2):2300471.

Buchfink, B., Xie, C., and Huson, D. H. (2015). Fast and sensitive protein alignment using DIAMOND. *Nature Methods*, 12(1):59–60.

Chen, Z., Zhao, P., Li, C., Li, F., Xiang, D., Chen, W., Zhang, T., Bai, Y., and Yang, M. (2024). DualNetGO: A Dual Network Model for Protein Function Prediction via Effective Feature Selection. *Bioinformatics*, 40(7):btae437.

Chua, Z. M., Rajesh, A., Sinha, S., and Adams, P. D. (2024). PROTGOAT: Improved Automated Protein Function Predictions Using Protein Language Models. *bioRxiv*, pages 1–15.

Elnaggar, A., Heinzinger, M., Dallago, C., Rehawi, G., Wang, Y., Jones, L., Gibbs, T., Feher, T., Angerer, C., Steinegger, M., Bhowmik, D., and Rost, B. (2021). ProtTrans: Toward Understanding the Language of Life Through Self-Supervised Learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(10):7112–7127.

Fan, K., Guan, Y., and Zhang, Y. (2020). Graph2GO: A Multi-Modal Attributed Network Embedding Method for Inferring Protein Functions. *GigaScience*, 9(8):giaa081.

Gillis, J. and Pavlidis, P. (2011). The impact of multifunctional genes on "guilt by association" analysis. *PLoS One*, 6(2):e17258.

Gligorijevic, V., Barot, M., and Bonneau, R. (2018). deepNF: Deep Network Fusion for Protein Function Prediction. *Bioinformatics*, 34(22):3873–3881.

He, K., Zhang, X., Ren, S., and Sun, J. (2016). Deep residual learning for image recognition. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 770–778.

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Zidek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S. A. A., Ballard, A. J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., Back, T., Petersen, S., Reiman, D., Clancy, E., Zielinski, M., Steinegger, M., Pacholska, M., Berghammer, T., Bodenstein, S., Silver, D., Vinyals, O., Senior, A. W., Kavukcuoglu, K., Kohli, P., and Hassabis, D. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature*, 596(7873):583–589.

Liu, Q., Zhang, C., and Freddolino, L. (2024). InterLabelGO+: unraveling label correlations in protein function prediction. *Bioinformatics*, 40(11):btae655.

Liu, Z., Mao, H., Wu, C.-Y., Feichtenhofer, C., Darrell, T., and Xie, S. (2022). A ConvNet for the 2020s. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 11976–11986.

Oliveira, G. B., Pedrini, H., and Dias, Z. (2024). Integrating transformers and automl for protein function prediction. In *46th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 1–5. IEEE.

Oliveira, G. B., Pedrini, H., and Dias, Z. (2025). SUPERMAGOv2: Protein Function Prediction via Transformer Embeddings and Bitscore-Weighted Features. *IEEE Access*, 13:139743–139757.

Piovesan, D., Giollo, M., Ferrari, C., and Tosatto, S. C. (2015). Protein function prediction using guilty by association from interaction networks. *Amino Acids*, 47(12):2583–2592.

Radivojac, P., Clark, W. T., Oron, T. R., Schnoes, A. M., Wittkop, T., Sokolov, A., Graim, K., Funk, C., Verspoor, K., Ben-Hur, A., Pandey, G., Yunes, J. M., Talwalkar, A. S., Repo, S., Souza, M. L., Piovesan, D., Casadio, R., Wang, Z., Cheng, J., Fang, H., Gough, J., Koskinen, P., Törönen, P., Nokso-Koivisto, J., Holm, L., Cozzetto, D., Buchan, D. W. A., Bryson, K., Jones, D. T., Limaye, B., Inamdar, H., Datta, A., Manjari, S. K., Joshi, R., Chitale, M., Kihara, D., Lisewski, A. M., Erdin, S., Venner, E., Lichtarge, O., Rentzsch, R., Yang, H., Romero, A. E., Bhat, P., Paccanaro, A., Hamp, T., Kaßner, R., Seemayer, S., Vicedo, E., Schaefer, C., Achten, D., Auer, F., Boehm, A., Braun, T., Hecht, M., Heron, M., Hönigschmid, P., Hopf, T. A., Kaufmann, S., Kiening, M., Krompass, D., Landerer, C., Mahlich, Y., Roos, M., Björne, J., Salakoski, T., Wong, A., Shatkay, H., Gatzmann, F., Sommer, I., Wass, M. N., Sternberg, M. J. E., Škunca, N., Supek, F., Bošnjak, M., Panov, P., Džeroski, S., Šmuc, T., Kourmpetis, Y. A. I., van Dijk, A. D. J., ter Braak, C. J. F., Zhou, Y., Gong, Q., Dong, X., Tian, W., Falda, M., Fontana, P., Lavezzo, E., Di Camillo, B., Toppo, S., Lan, L., Djuric, N., Guo, Y., Vucetic, S., Baumgartner, A., Weskamp, N., Kramer, S., Lassmann, T., Ren, J., Wong, W. H., and Friedberg, I. (2013). A large-scale evaluation of computational protein function prediction. *Nature Methods*, 10(3):221–227.

Rives, A., Meier, J., Sercu, T., Goyal, S., Lin, Z., Liu, J., Guo, D., Ott, M., Zitnick, C. L., Ma, J., and Fergus, R. (2021). Biological structure and function emerge from scaling unsupervised learning to 250 million protein sequences. *Proceedings of the National Academy of Sciences*, 118(15):e2016239118.

Satyanarayana, U. (2013). *Biochemistry*. Elsevier India.

Su, J., Zhu, M., Murtadha, A., Pan, S., Wen, B., and Liu, Y. (2022). ZLPR: A Novel Loss for Multi-label Classification. *arXiv:2208.02955*, pages 1–17.

Suzek, B. E., Huang, H., McGarvey, P., Mazumder, R., and Wu, C. H. (2007). UniRef: comprehensive and non-redundant UniProt reference clusters. *Bioinformatics*, 23(10):1282–1288.

Szklarczyk, D., Gable, A. L., Lyon, D., Junge, A., Wyder, S., Huerta-Cepas, J., Simonovic, M., Doncheva, N. T., Morris, J. H., Bork, P., Jensen, L. J., and Mering, C. v. (2021). The STRING Database in 2021: Customizable Protein-Protein Networks, and Functional Characterization of User-Uploaded Gene/Measurement Sets. *Nucleic Acids Research*, 49(D1):D605–D612.

The UniProt Consortium (2024). UniProt: the Universal Protein Knowledgebase in 2025. *Nucleic Acids Research*, 53(D1):D609–D617.

Ulusoy, E. and Dogan, T. (2025). ProtHGT: Heterogeneous Graph Transformers for Automated Protein Function Prediction Using Biological Knowledge Graphs and Language Models. *bioRxiv*, pages 1–29.

Valentini, G. (2010). True Path Rule Hierarchical Ensembles for Genome-Wide Gene Function Prediction. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 8(3):832–847.

Wang, Y., Zhang, Y., Pu, M., Zhang, K., Chen, Y., Li, M., and Wu, F.-X. (2025). SEGT-GO: a graph transformer method based on PPI serialization and explanatory artificial intelligence for protein function prediction. *BMC Bioinformatics*, 26(1):46.

Xu, X., Deng, Y., Bronswijk, C., and Bonvin, A. M. J. J. (2024). DeepRank-GNN-esm: a graph neural network for scoring protein–protein models using protein language model. *Bioinformatics Advances*, 4(1):vbad191.

You, R., Yao, S., Xiong, Y., Huang, X., Sun, F., Mamitsuka, H., and Zhu, S. (2018). DeepGO: predicting protein functions from sequence and interactions using a deep ontology-aware classifier. *Bioinformatics*, 34(4):660–668.

You, R., Yao, S., Xiong, Y., Huang, X., Sun, F., Mamitsuka, H., and Zhu, S. (2021). DeepGraphGO: Graph Neural Network for Large-Scale, Multispecies Protein Function Prediction. *Bioinformatics*, 37:i262–i271.

Zhong, G., Chang, X., Xie, W., and Zhou, X. (2024). Targeted Protein Degradation: advances in drug discovery and clinical practice. *Signal Transduction and Targeted Therapy*, 9(1):308.

Zhou, N., Jiang, Y., Bergquist, T. R., Lee, A. J., Kacsoh, B. Z., Crocker, A. W., Lewis, K. A., Georghiou, G., Nguyen, H. N., Hamid, M. N., Davis, L., Dogan, T., Atalay, V., Rifaioglu, A. S., Dalkiran, A., Cetin Atalay, R., Zhang, C., Hurto, R. L., Freddolino, P. L., Zhang, Y., Bhat, P., Supek, F., Fernández, J. M., Gemovic, B., Perovic, V. R., Davidovic, R. S., Sumonja, N., Veljkovic, N., Asgari, E., Mofrad, M. R. K., Profiti, G., Savojardo, C., Martelli, P. L., Casadio, R., Boecker, F., Schoof, H., Kahanda, I., Thurlby, N., McHardy, A. C., Renaux, A., Saidi, R., Gough, J., Freitas, A. A., Antczak, M., Fabris, F., Wass, M. N., Hou, J., Cheng, J., Wang, Z., Romero, A. E., Paccanaro, A., Yang, H., Goldberg, T., Zhao, C., Holm, L., Törönen, P., Medlar, A. J., Zosa, E., Borukhov, I., Novikov, I., Wilkins, A., Lichtarge, O., Chi, P.-H., Tseng, W.-C., Linial, M., Rose, P. W., Dessimoz, C., Vidulin, V., Dzeroski, S., Sillitoe, I., Das, S., Lees, J. G., Jones, D. T., Wan, C., Cozzetto, D., Fa, R., Torres, M., Warwick Vesztrocy, A., Rodriguez, J. M., Tress, M. L., Frasca, M., Notaro, M., Grossi, G., Petrini, A., Re, M., Valentini, G., Mesiti, M., Roche, D. B., Reeb, J., Ritchie, D. W., Aridhi, S., Alborzi, S. Z., Devignes, M.-D., Koo, D. C. E., Bonneau, R., Gligorijevic, V., Barot, M., Fang, H., Toppo, S., Lavezzo, E., Falda, M., Berselli, M., Tosatto, S. C. E., Carraro, M., Piovesan, D., Rehavi, H., Kaplan, S., Mazor, S., Nahum, T., Szilágyi, G., Doğan, T., Smuc, T., Supek, F., Gough, J., Orengo, C., Hamp, T., Kassner, R., Seemayer, S., Vicedo, E., Schaefer, C., Achten, D., Auer, F., Boehm, A., Braun, T., Hecht, M., Heron, M., Hönigschmid, P., Hopf, T. A., Kaufmann, S., Kiening, M., Krompass, D., Landerer, C., Mahlich, Y., Roos, M., Björne, J., Salakoski, T., Wong, A., Shatkay, H., Gatzmann, F., Sommer, I., Wass, M. N., Sternberg, M. J. E., Škunca, N., Šuperti Furga, G., Lancet, D., Petrov, D., Paladin, L., Peterlongo, P., Tiengo, A., Gemovic, B., Perovic, V. R., Veljkovic, N., Veljkovic, N., Moesa, H. A., Bromberg, Y., Zhu, S., Wang, J., Cao, R., Cheng, J., Altenhoff, A., Rives, A., Kulmanov, M., Georghiou, G., Götz, S., Forslund, K., Sonnhammer, E. L. L., Cozzetto, D., Messih, M. A., Törönen, P., Ghebremariam, Y. B., Gromiha, M. M., Kihara, D., Oruganty, K., Kryshtafovych, A., Fidelis, K., Pazos, F., Tress, M. L., Valencia, A., Blanc, B., George, R. A., Grishin, N. V., Radivojac, P., and Friedberg, I. (2019). The CAFA Challenge Reports Improved Protein Function Prediction and New Functional Annotations for Hundreds of Genes Through Experimental Screens. *Genome Biology*, 20(1):1–23.