Michael Vaden

Georgia Institute of Technology, Atlanta, USA

Graph of Experts: Trainable Graphs for Compositional Reasoning

put abstract here

0 Introduction

-hierarchical problems require systems that adapt across levels of abstraction

-vesuvius scrolls interesting because of diversity of challenges

-conventional deep networks such as UNets or transformers apply the same operations to every region, lacking mechanisms to dynamically compose specialized behaviors

-reframe modular neural routing as a learnable graph

-each node goes with an expert trained for distinct subtask

-in the vesuvius case, geometry reconstruction, fiber orientation, ink segmentation

-directed nodes encode transition policies governing how information propagates

-GoE learns structured pathways through the expert graph, discovering intermediate representations to solve complex, multi-scale problems

-unifies specialization and coordination

-gating mechanism dynamically selects experts and edge transitions

-allows for self-organization into meaningful workflows

-Add experiment info

1 Background and Related Work

-Recent progress in adaptive architectures explores dividing neural computation into specialized components

-MoE paradigm trains multiple experts in a specific data regime, while a gating network selects which to activate per input

-Improves efficiency but remains flat due to one pass

-Google's mixture of recursions introduces recursive expert calls, anbling dynamic reasoning chains, these are typically linear or sequential, optimized for symbolic or temporal reasoning tasks

-They lack explicit modeling of relationships between experts themselves, for instance, how information should transition between specialists handling distinct subproblems

-Each edge encodes transition policies

-Gating mechanism operates not as a simple router but as a graph traversal policy

-Trained to discover efficient and semantically meaningful pathways across experts

-Learns compositional workflows rather than isolated specializations

-Mention vesuvius relation

2 Graph of Experts

In this work we introduce Graph of Experts (GoE) as a novel architecture for compositional and structured reasoning across interconnected neural specialists. Unlike traditional mixture-of-experts models that operate in a single routing step, GoE organizes experts as nodes within a directed graph whose edges encode learnable transition policies. Each expert learns a specialized transformation, while the graph topology and routing dynamics determine how information flows and recombines across experts.

2.1 Problem

We consider GoE as a unified model for the coupled tasks presented by the Vesuvius Challenge of geometry unwrapping and ink detection.

2.2 Input Parameterization

The input stem serves as the interface between the raw volumetric scans of the Vesuvius scrolls and the Graph of Experts model, converting high-dimensional tomographic data into standardized latent and tokenized representations suitable for downstream routing and reasoning. Each input volume $x \in \mathbb{R}^{B \times C \times D \times H \times W}$ represents a 3D X-ray segment containing papyrus fibers, voids, and potential ink traces.

A 3D convolutional encoder $\phi(x; \theta_s)$ extracts localized features like carbonization gradients, fiber directionality, and ink-related density anomalies—across the scroll's depth. To enable compositional reasoning, f_0 is partitioned into overlapping voxel patches corresponding to localized surface regions of the papyrus. Each patch ω_i is projected into a shared embedding space via a learned projection W_t, producing a set of token embeddings

$$t_i = W_t \psi(f_0, \Omega_i) + b_t, \qquad T = \{t_1, \ldots, t_N\} \tag{1}$$

where ψ denotes the patch extraction operation. These tokens $T \in \mathbb{R}^{B \times N \times d}$ form a modality-independent representation that encodes spatial and material context across the scroll's interior.

In addition to token embeddings, the input stem computes a compact set of auxiliary routing features a_i that describe local physical structure within each patch

$$a_i = g_{\text{aux}}(f_0[\Omega_i]) \tag{2}$$

Here g_{aux} is implemented as a lightweight 3D convolutional projection capturing curvature, fiber orientation, and anisotropy statistics. These features provide the graph router with cues about surface geometry and structural continuity, guiding tokens toward experts specialized in geometry unwrapping, fiber estimation, or ink segmentation.

2.3 Encoder

After the input stem extracts token embeddings and auxiliary routing features from volumetric data, the encoder models global dependencies between these spatially localized representations. Its purpose is to capture long-range relationships across the scroll—linking ink patterns, papyrus curvature, and fiber structures that may be spatially distant yet physically or semantically related. Formally, the encoder applies a series of transformer-based self-attention blocks to the token set

$$T' = \Phi(T; \theta_e) \tag{3}$$

where each layer follows the sequence LayerNorm → Multi-Head Self-Attention → Residual → FeedForward → Residual. The self-attention mechanism allows the model to reason across the unwrapped 3D surface, dynamically weighting interactions between regions that share similar geometric or material signatures. For example, two distant patches may attend strongly if they belong to the same continuous ink stroke revealed through unwrapping.

Each attention head computes

$$\text{Attn}(Q, K, V) = \text{softmax}\left(\frac{QK^\top}{\sqrt{d_k}}\right)V \tag{4}$$

enabling the encoder to infer relationships such as fiber alignment continuity, cross-layer ink correlation, and curvature transitions that traditional convolutional methods may miss.

In parallel, the auxiliary routing features are refined to remain aligned with the evolving contextual embeddings

$$a' = f_{\text{aux}}(a, T') \tag{5}$$

where f_{aux} applies a lightweight projection or cross-attention between the routing features and the updated token embeddings. This ensures that routing decisions later in the graph are influenced not only by local texture cues but also by global geometric context.

By operating in token space, the encoder effectively performs contextual reasoning across the 3D volume, transforming purely local voxel representations into globally informed embeddings. The resulting pair (T', a') forms a unified, context-aware representation of the scroll, ready for graph-based expert routing — where specialized nodes handle geometry reconstruction, fiber orientation, and ink detection in a coordinated manner.

2.4 Graph Router

The graph router governs how information flows between specialized experts that handle distinct aspects of the Vesuvius scroll problem—geometry reconstruction, fiber orientation, and ink segmentation. Rather than treating routing as a single gating step, the router defines a learnable directed graph $G = (V, E)$ where each node $v_m \in V$ represents a domain-specific expert, and the edges encode adaptive dependencies between them.

For each input token t_i' and its routing descriptor a_i', a combined representation $r_i = [t_i'; a_i']$ is formed and mapped to a probability distribution over experts

$$a' = f_{\text{aux}}(a, T') \tag{6}$$

where $p_{i,m}$ indicates the strength with which token i is routed to expert m. Tokens thus enter the graph through different entry points, reflecting their local characteristics—flat papyrus regions may flow toward geometry experts, while high-density or anisotropic patches route toward ink specialists. Once inside the graph, tokens propagate along directed edges weighted by A_{mn}, which represent learned communication channels between related experts. For instance, geometry and fiber experts may exchange contextual information before passing refined representations to the nk expert. Each node updates its state as

$$h_m^{l+1} = \rho\Big(\sum_{n \in N(m)} A_{mn} W_n h_n^l\Big) \tag{7}$$

where ρ is a nonlinear activation and $N(m)$ denotes the neighbors of node m.

In the Vesuvius context, this structure captures the natural dependency hierarchy of the problem: geometry precedes unwrapping, fibers align with curvature, and ink must be detected on the reconstructed surface. By learning transition policies A_{mn}, the router discovers efficient and interpretable processing chains that mirror the underlying physical relationships between these tasks. Routing thus follows three principles:

1. *Auxiliary features* bias entry points, meaning tokens with similar auxiliary cues cluster to the same experts.
2. *Graph topology* structures the reasoning process in how information flows through connected specialists.
3. *Dynamic routing* allows probabilites to determine to most relevant expert pathways. specialists.

This approach transforms the Vesuvius reconstruction process into a trainable workflow graph, allowing the system to jointly reason about geometry, structure, and ink through adaptive expert coordination rather than static pipelines.

2.5 Experts and Refinement

Each node $v_m \in V$ of the expert graph corresponds to a learned specialist, not pre-assigned to any predefined subtask. During training, experts self-organize into functional roles through the routing dynamics of the model—some may implicitly focus on geometry-like structure, others on fiber consistency or ink-related features—but these specializations arise entirely from data-driven pressure rather than architectural constraints.

Formally, each expert implements a transformation

$$h_m^{l+1} = f_0(h_m^l; \theta_m) \tag{8}$$

where h_m^l represents the set of token embeddings routed to expert m at iteration l, and f_m is a general neural operator such as a transformer block. The model imposes no prior assumptions about the spatial or semantic meaning of any expert; instead, the routing process determines which features each expert learns to refine.

Token embeddings are aggregated into expert states according to their routing probabilities

$$\bar{h}_m = \sum_i p_{i,m} W_p r_i \tag{9}$$

ensuring that each expert receives a differentiable mixture of the tokens most relevant to its learned domain of competence. Experts exchange information through directed edges in the graph, which define learnable communication channels

$$h_m^{l+1} = f_0(\bar{h}_m; \theta_m) + \sum_i A_{mn} W_c h_n^l \tag{10}$$

allowing representations to evolve cooperatively across connected specialists.

To maintain balanced usage across experts, the router is regularized using load-balancing and entropy terms

$$L_{\text{balance}} = \lambda_b\left(\frac{1}{B}\sum_i p_{i,m} - \frac{1}{M}\right)^2 \tag{11}$$

$$L_{\text{entropy}} = -\lambda_e\left(\sum_i \sum_m p_{i,m} \log p_{i,m}\right) \tag{12}$$

which prevent the network from collapsing onto a single dominant node and encourage the emergence of diverse, interpretable subspaces.

The model employs an iterative refinement cycle in which experts repeatedly reprocess and exchange information

$$H^{k+1} = F(H^k; \theta_{\text{exp}}; \theta_{\text{router}}), \quad k = 0, \ldots, K+1 \tag{13}$$

where $H_k = \{h_1^k, \ldots, h_M^k\}$ denotes the collective state of all experts. At each iteration:

1. *Re-routing:* Updated token embeddings are re-scored by the router, allowing adaptive reassignment of ambiguous tokens.
2. *Expert Update:* Each expert processes its aggregated inputs and incorporates messages from neighboring experts.
3. *Convergence Check:* Confidence heads estimate token-level stability, halting recursion when representations converge or uncertainty falls below a threshold.

This recursive refinement process enables experts to self-organize into functional hierarchies without explicit supervision. In the Vesuvius context, this means that the model can autonomously discover experts that align with distinct phenomena—such as surface geometry, fiber continuity, or ink texture—purely through exposure to volumetric data and reconstruction objectives. The result is a flexible, compositional reasoning system in which specialization and coordination both emerge from learned structure, allowing the network to adaptively allocate computation across different aspects of the scroll's complex 3D representation.

2.6 Decoders

The decoder projects the refined expert representations back into the spatial domain, translating the model's internal reasoning into interpretable outputs suitable for the Vesuvius scroll reconstruction task. Rather than receiving explicitly labeled expert outputs, the decoder integrates the final states of all learned specialists—each having converged toward a self-organized role through training—and reconstructs the underlying volumetric structures and material contrasts of the papyrus.

After the iterative refinement process, the final expert states $H^K = \{h_1^K, \ldots, h_M^K\}$ are aggregated according to their routing distributions

$$z_i = \sum_m p_{i,m} W_d h_m^K \tag{14}$$

producing producing a unified latent embedding z_i for each token. These embeddings are spatially reassembled into the 3D coordinate grid using positional information from the input stem, forming a dense reconstruction volume.

The decoder's architecture remains deliberately modality-agnostic—it learns how to map latent embeddings back to the desired supervision targets (e.g., surface geometry, unwrapped intensity, or ink probability) without assuming explicit expert-task pairings. Each output head is trained to predict a particular modality through learned attention over the aggregated representation

$$y_k = D_k(Z; \theta_{D_k}), \quad k \in \{1, \ldots, K_{\text{out4}}\} \tag{15}$$

where each D_k produces a spatial prediction such as reconstructed geometry, fiber orientation field, or ink density map. The correspondence between experts and heads is discovered during optimization, driven by reconstruction loss gradients rather than predefined structure.

Supervision is provided via multi-objective loss terms reflecting the available ground truth or proxy targets

$$L_{\text{total}} = \sum_k \lambda_k L_k(y_k, y_k^*) \tag{16}$$

where y_k^* are reference maps derived from unwrapped surfaces, volumetric annotations, or ink detection benchmarks. Each loss contributes to refining the shared latent space so that specialized experts implicitly align with distinct predictive tasks.

In the Vesuvius domain, this enables the model to reconstruct geometry and ink jointly, with the decoder learning to interpret whatever representations emerge from the expert graph. The result is a fully learned end-to-end system that converts raw 3D tomography into structured, interpretable outputs—recovering the physical and textual layers of the scroll without manual decomposition into subtasks.

3 Experiments