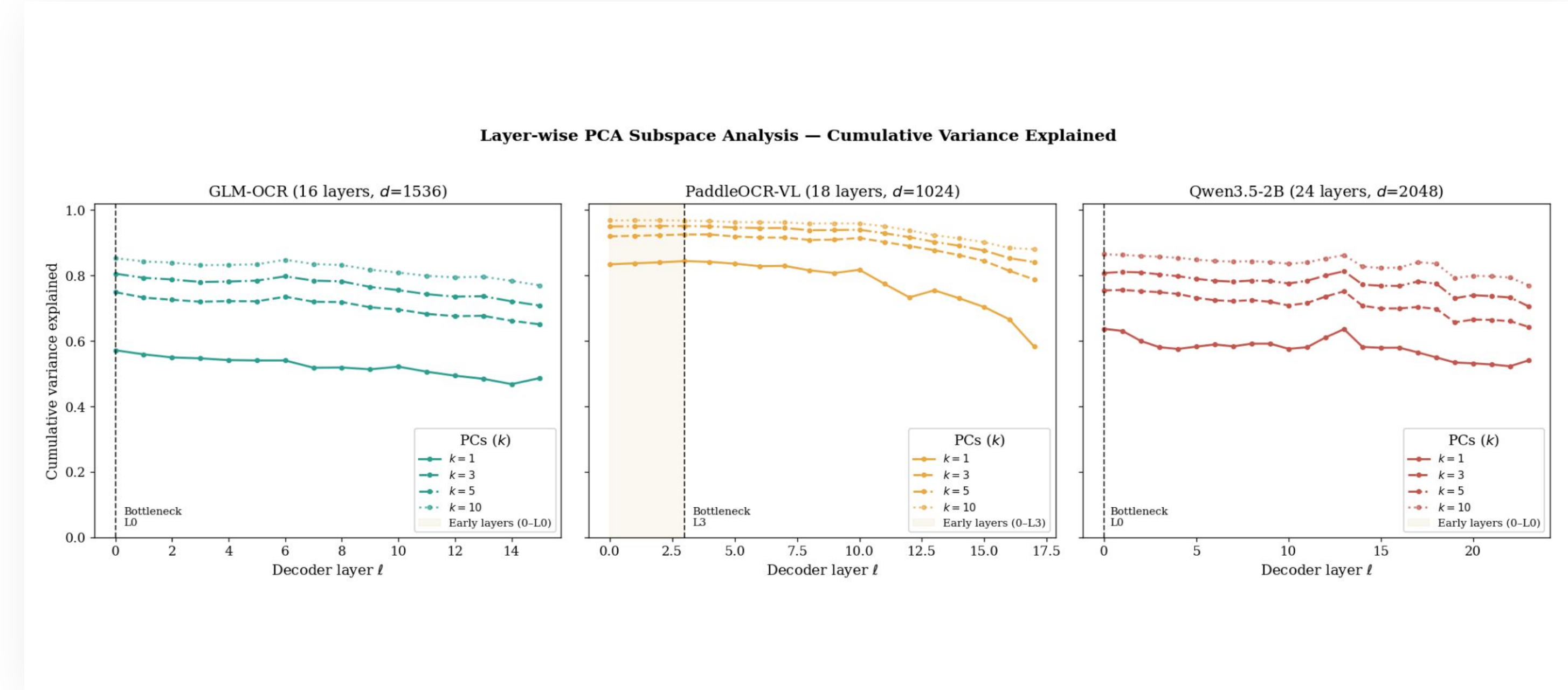


OCR models exhibit concentrated, low-dim document structure subspaces compared to general-purpose vision-language models, despite high cross-architecture alignment.

Low-Dimensional Document Structure Subspaces in Specialized vs. Emergent OCR Models: An Interpretability Study of Three Architectures

Background: Optical character recognition (OCR) models are critical infrastructure, but their internal representations remain unexplored. We compared those with general-purpose document recognition systems.

Result 1: Extreme Representational Compression
All three models concentrate OCR information into low-dimensional subspaces at early decoder depths, especially the specialized OCR models.



Result 2: Document Structure Modularity Index: M
 M formalizes the representational organization. Both specialized models achieve higher M than the baseline, showing that representations for text, tables, and mixed content occupy partially disentangled subspaces.

Model	GLM	Paddle	Qwen
M	.774	.715	.704

Methods

1 Extract Internal Decoder Representations

GLM-OCR (0.9B to 2.0B params)
PaddleOCR-VL (0.9B to 2.0B params)
Qwen3.5 (0.8B to 2.0B params)

Activations were hooked from three architectures (GLM-OCR, PaddleOCR-VL, Qwen3.5) ranging from 0.9B to 2.0B parameters.

2 PCA Analysis of 300 Documents

100 TEXT, 100 TABLE, 100 MIXED CONTENT

Principal Component Analysis was performed on 100 text, 100 table, and 100 mixed-content images.

3 Measure Alignment and Modularity

ALIGNMENT & MODULARITY

- Grassmann distances
- principal angles
- Cross-Model CKA

Subspaces were compared using Grassmann distances, principal angles, and Cross-Model CKA to determine representational similarity.

Limitation: Our projection interventions are activation ablations rather than full causal tracing, establishing that identified subspaces are necessary for OCR performance, but not proving sufficiency.