

MatCurvs: Article Real-Coordinate Curve Extraction for Agent-Ready Materials Reasoning

Liang Yin^{1,2}, Zhan'ao Yao¹, Jiahui Shi^{3,4,*}, Songlin Yu^{1,4,*}, Jianjun Liu^{1,2,5*}

¹ Shanghai Institute of Ceramics, CAS, ² School of Advanced Interdisciplinary Sciences, UCAS, ³ Institute of Automation, CAS, ⁴ MatTok (Shanghai) AI Technology Co., Ltd., ⁵ Hangzhou Institute for Advanced Study, UCAS

1. Why Materials Reasoning Needs Numerical Data

Materials-science figures contain evidence needed to compare materials and estimate physical properties. XRD peak positions and widths, Raman D/G ratios, and EIS intercepts all depend on precise numerical measurements.

Current VLMs often describe the overall trend but miss detailed per-curve coordinates. MatDeplot turns published curves into reusable numerical data, as illustrated in Figure 1.

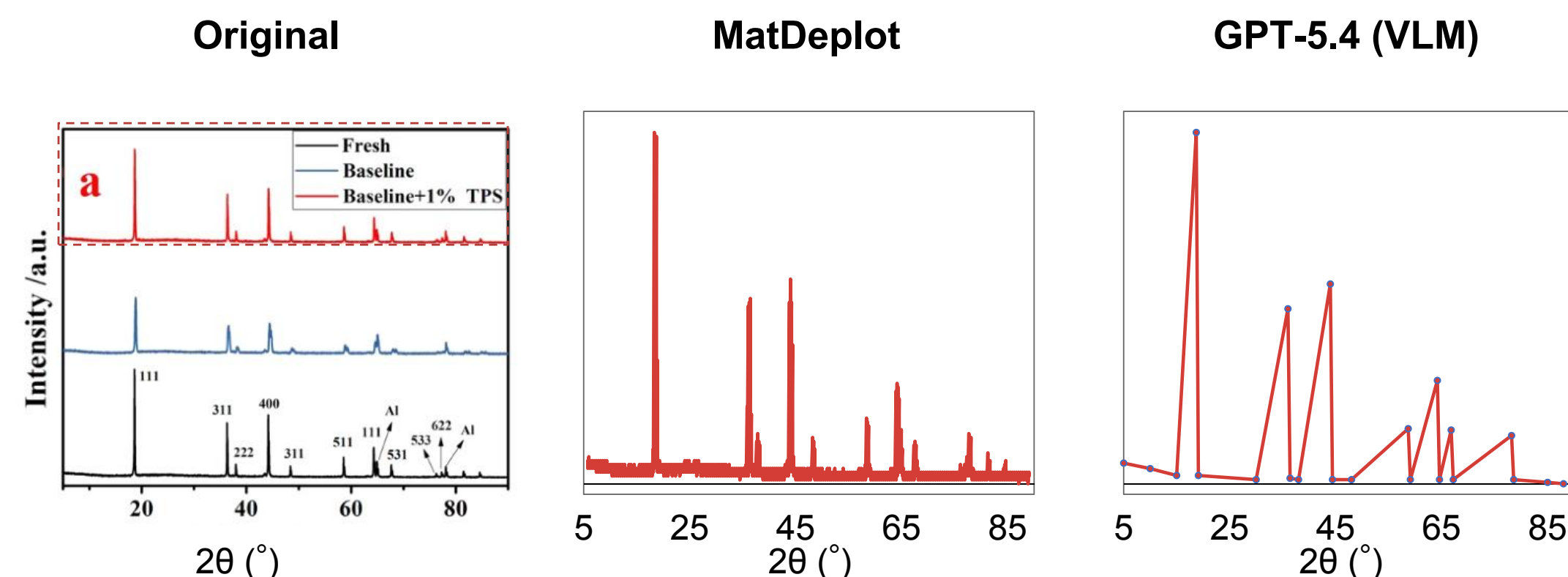


Figure 1. Reconstruction of the highlighted top XRD trace. MatDeplot retains narrow peaks; GPT-5.4 returns a coarse 22-point approximation. Shared x-axis; intensity normalized per trace.

Complete trajectories let an agent query, integrate and compare published measurements using explicit numerical operations.

2. MatDeplot: Recovering Numerical Curves from Figures

MatDeplot connects document parsing, subfigure routing, axis calibration and curve extraction (Figure 2). It produces a calibrated data-space trajectory for each detected curve.

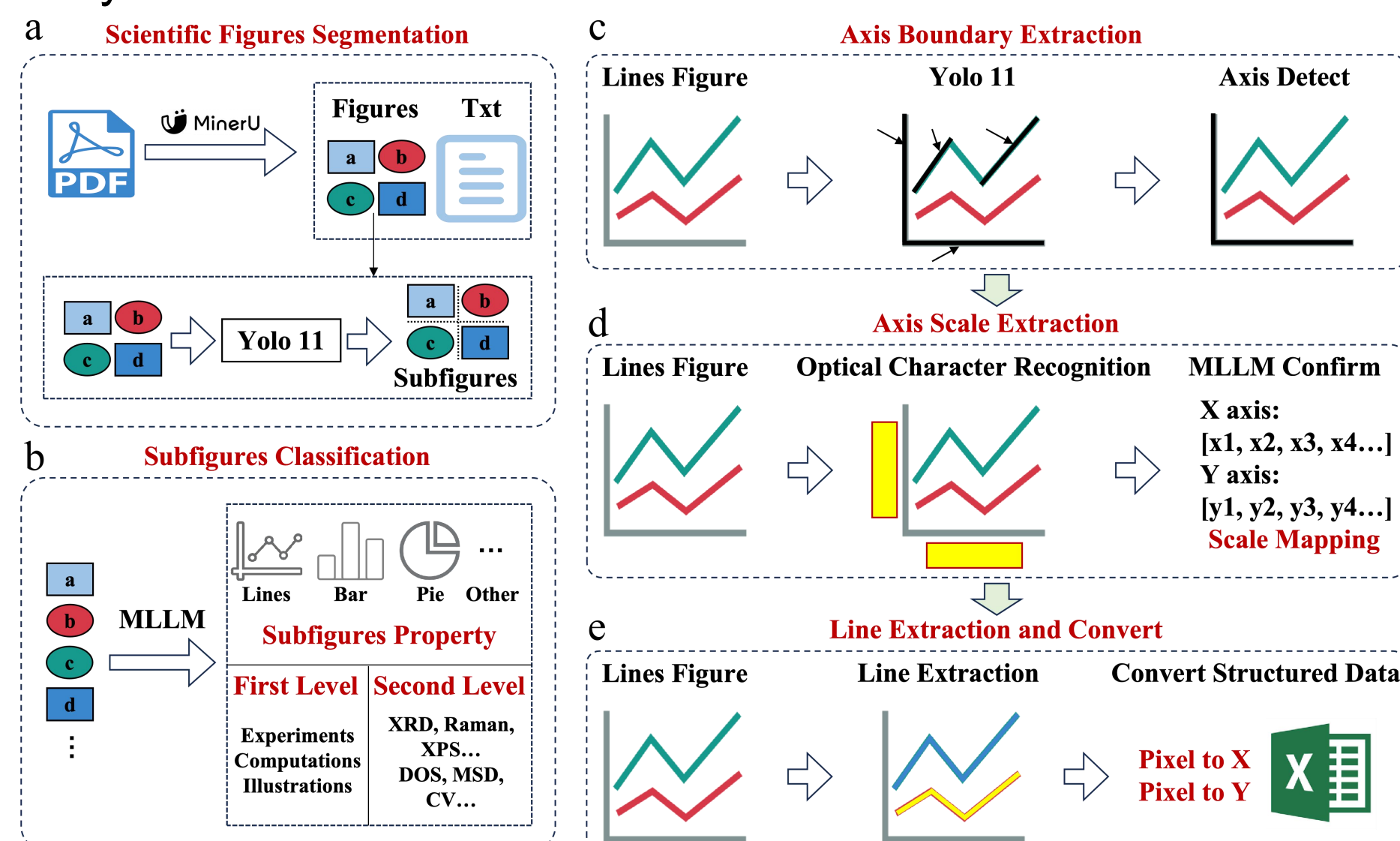


Figure 2. The article-to-curve pipeline: extract and classify subfigures, locate axes, read scales, recover curves and export numerical values.

A Mask2Former-style head on MiT-B3 isolates foreground pixels^[1, 2]. The reader separates curve instances using both color and spatial position, then calibrates their numerical coordinates.

$$\text{Pixel features} \quad \varphi(u, v) = [L^*, a^*, b^*, 0.5u', 0.5v']$$

LAB color and standardized image position determine the clusters. Each cluster is skeletonized into a pixel polyline. Numeric axis ticks determine the mapping from image positions to data-space values:

$$\text{Axis calibration} \quad x = a_x u + b_x, \quad y = a_y v + b_y$$

Outputs retain curve identity, pixel polylines and calibrated samples for downstream lookup, interpolation, integration and comparison.

3. MatCurvs: Real-Coordinate Curves for Materials Reasoning

MatCurvs links literature-scale discovery, manual pixel annotation and real-coordinate measurements. The three tiers separate corpus routing, geometric evaluation and downstream numerical use.

Table 1. The three evidence levels in MatCurvs.

Tier	Scale	Annotation and role
L0	1,375,165 subfigures	Corpus routing; 55,763 articles
L1	204 panels (90 clean)	Manual per-curve pixel polylines
L2	8,026 curves / 3,817 panels	Calibrated real-coordinate series

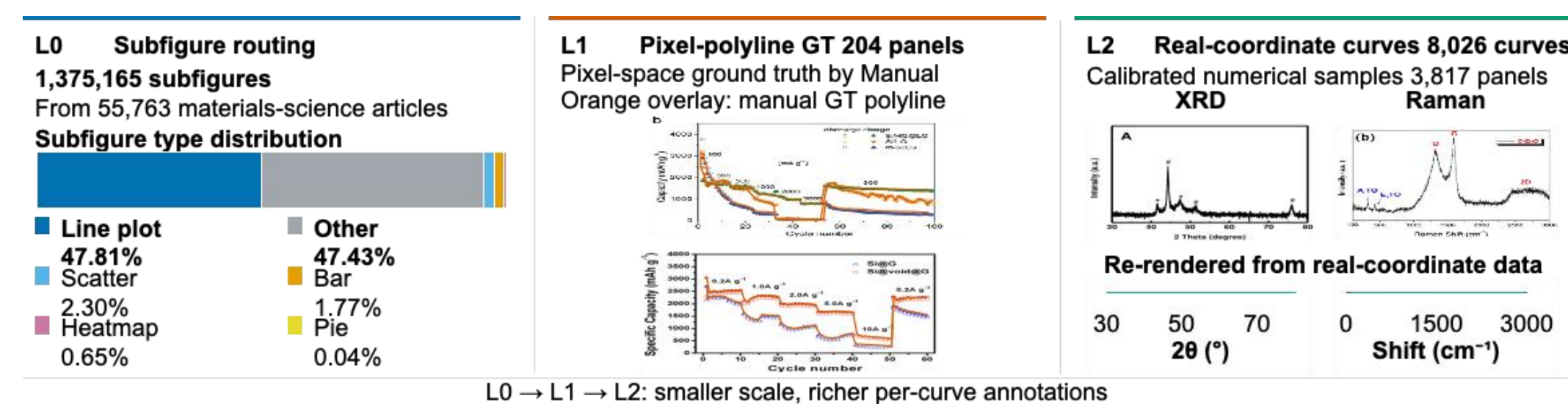


Figure 3. From left to right: routed figure types, manual curve annotations, and original figures paired with extracted real-coordinate curves.

Seven families: XRD, Raman, XPS, XAFS, EIS, GCD and CV. MatCurvs-Reasoning provides 14,740 numerical questions.

4. Theoretical Analysis: Curve Reconstruction Fidelity

We evaluate whether recovered curves align with the original pixels and retain usable numerical coordinates.

$$\text{Pixel overlap} \quad \text{IoU}(P, G) = |P \cap G| / |P \cup G|$$

$$\text{Curve recall} \quad \text{LR}_\theta = \text{TP}_\theta / \text{NGT}$$

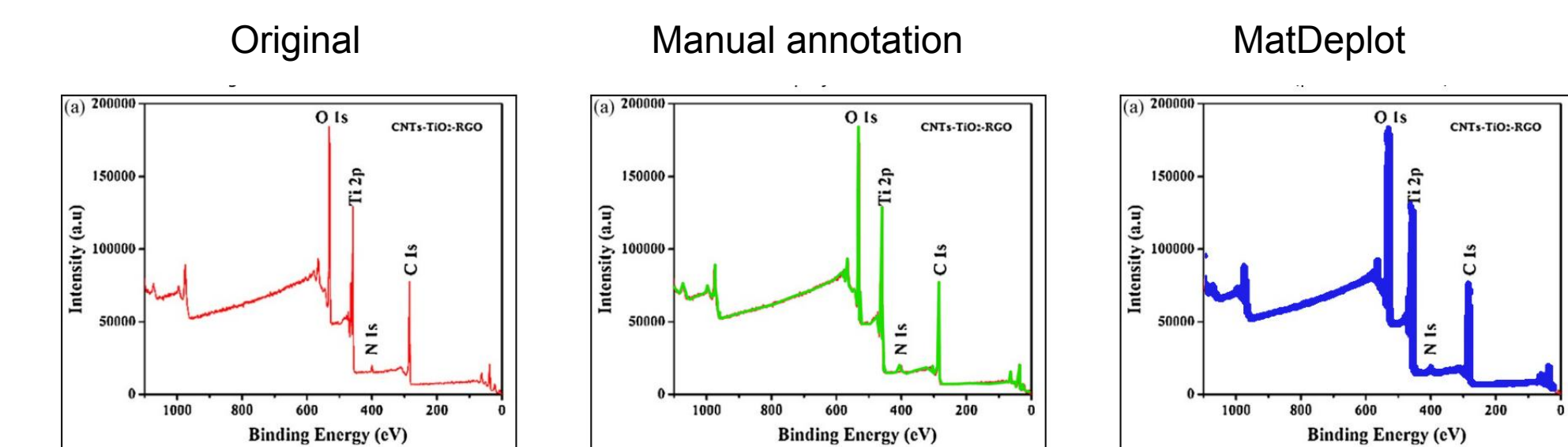
TP_θ : matched GT curves with $\text{IoU} \geq \theta$. NGT : all GT curves.

Point support is measured within ± 1 image column. Vertical error $|\hat{y} - y|$ is reported at supported locations.

5-pixel GT rasterization; one-to-one Hungarian matching.

5. Experiments: Extraction Accuracy and Reasoning Performance

Pixel-space reconstruction



Representative L1 reconstruction: original, GT and extracted curves.

MatDeplot 46% MatDeplot, 46%

Hosted VLMs 1% Hosted VLMs, 1%

0% 10% 20% 30% 40% 50%

LR0.5: 90 panels / 203 curves (VLM upper bounds).

Point support: 85.4%. Supported median vertical error: 2 px.

Table 2. Reasoning results: 49 panels / 188 questions

Method	Model	N	medRel ↓	Pass@5% ↑
M1 image only	qwen-vl-max	188	32.6%	21.3%
	gemini-3.1-pro-preview	188	26.4%	32.4%
	gpt-5.4	188	15.4%	37.2%
M2 image+data	qwen-vl-max	188	15.9%	28.2%
	gemini-3.1-pro-preview	188	4.7%	46.8%
	gpt-5.4	188	4.3%	53.2%
M3 data only	qwen3-max	188	14.0%	38.3%
	claude-opus-4.6	188	8.5%	44.1%
	gpt-5.4	188	4.6%	52.1%
M4 SciPy	—	188	6.7%	44.7%

Fixed-model reasoning

GPT-5.4; 87 questions from 26 single-curve panels.

Image only 19.4%

Image + data 4.3%

0% 5% 10% 15% 20% 25%

Median relative error ↓. Pass@5%: 32.2% → 54.0%.

References :

[1] Cheng et al. Mask2Former. CVPR, 2022.

[2] Xie et al. SegFormer. NeurIPS, 2021.

[3] Liu et al. DePlot. Findings of ACL, 2023.

