实验结果

基准评测结果

DeepSeek-OCR 在 Fox Benchmark 和 OmniDocBench 上的全面评估，展示了卓越的压缩效率和精度保持能力。

核心指标

97%
OCR 解码精度
压缩率 < 10× 时

60%
高压缩率精度
20× 压缩率时

100
视觉词元数
Small 模式

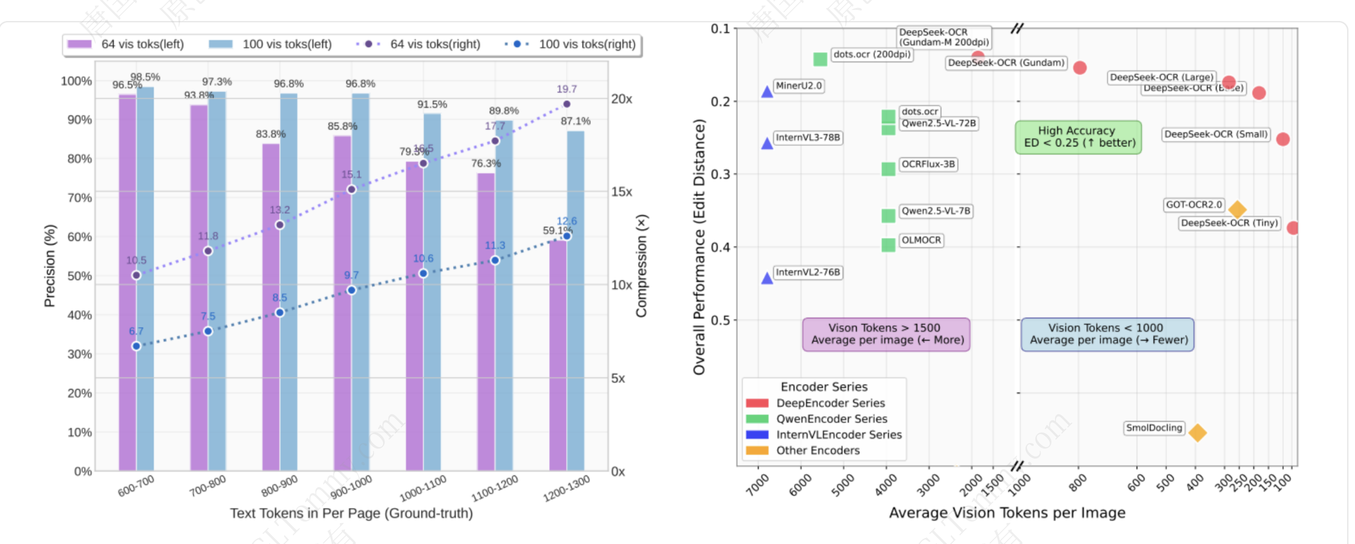
20万+
日处理页数
单 A100-40G GPU

Fox Benchmark 结果

视觉-文本压缩能力验证

测试不同压缩率下的 OCR 解码精度，验证光学压缩范式的有效性

图 1(a): 压缩率与精度关系



(a) Compression on Fox benchmark (b) Performance on Omnidocbench

Figure 1 | Figure (a) shows the compression ratio (number of text tokens in ground truth/number of vision tokens model used) testing on Fox [21] benchmark; Figure (b) shows performance comparisons on OmniDocBench [27]. DeepSeek-OCR can achieve state-of-the-art performance among end-to-end models enjoying the fewest vision tokens.

表 2: 不同压缩率下的精度数据

表 2: 压缩率与精度数据图表

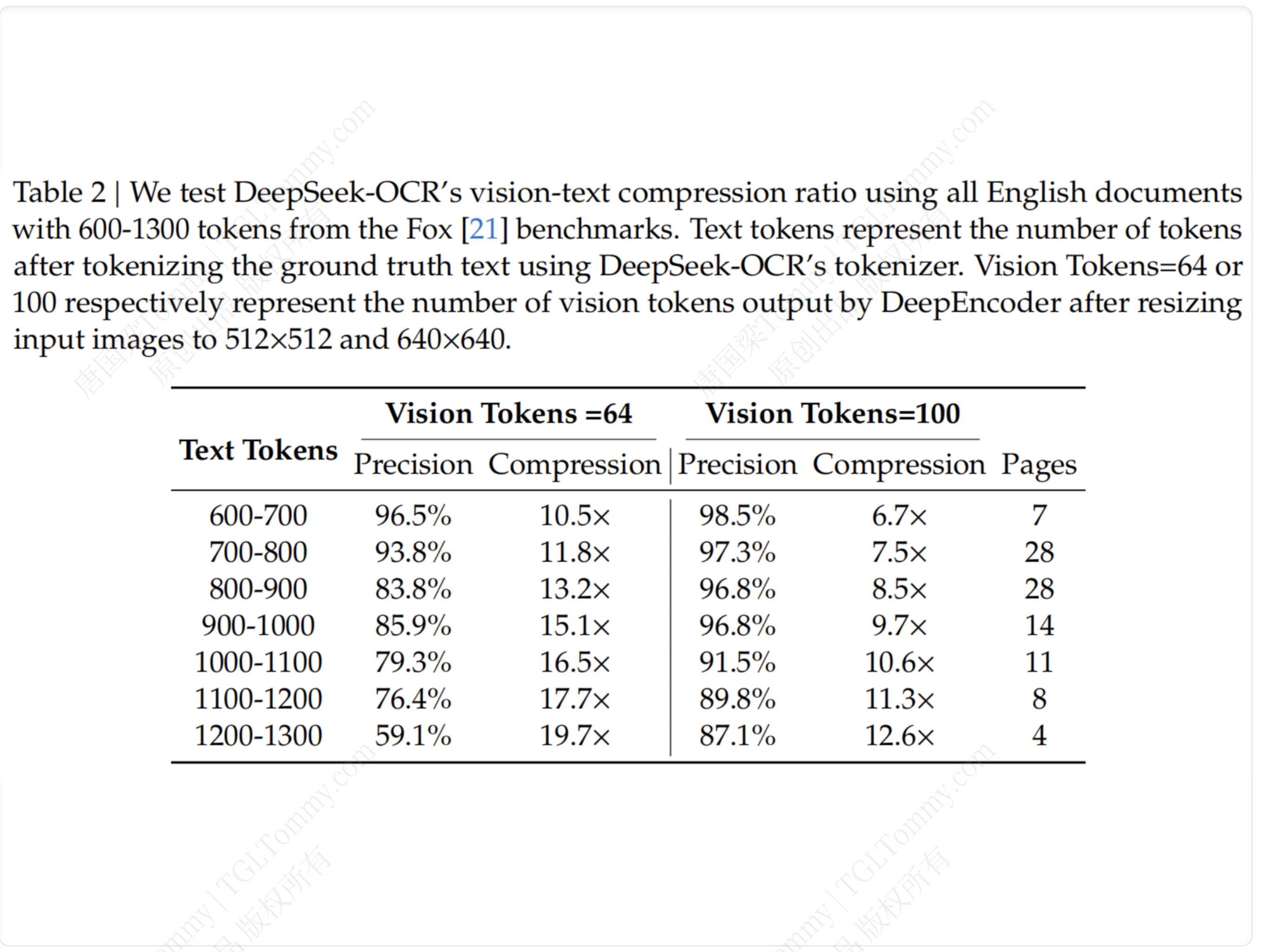


Table 2 | We test DeepSeek-OCR's vision-text compression ratio using all English documents with 600-1300 tokens from the Fox [21] benchmarks. Text tokens represent the number of tokens after tokenizing the ground truth text using DeepSeek-OCR's tokenizer. Vision Tokens=64 or 100 respectively represent the number of vision tokens output by DeepEncoder after resizing input images to 512×512 and 640×640.

Text Tokens	Vision Tokens =64		Vision Tokens=100		
	Precision	Compression	Precision	Compression	Pages
600-700	96.5%	10.5×	98.5%	6.7×	7
700-800	93.8%	11.8×	97.3%	7.5×	28
800-900	83.8%	13.2×	96.8%	8.5×	28
900-1000	85.9%	15.1×	96.8%	9.7×	14
1000-1100	79.3%	16.5×	91.5%	10.6×	11
1100-1200	76.4%	17.7×	89.8%	11.3×	8
1200-1300	59.1%	19.7×	87.1%	12.6×	4

关键发现

- 当压缩率在 10 倍以内时，模型可实现 **97% 的 OCR 解码精度**
- 即使在 20 倍压缩率下，精度仍能保持在 **60% 左右**
- 直接验证了**光学压缩范式的有效性**，证明视觉模态可作为文本信息的高效压缩介质

OmniDocBench 结果

OCR 实用性能验证

在真实文档 OCR 任务上的性能表现，与主流模型对比

表 3: OmniDocBench 性能对比

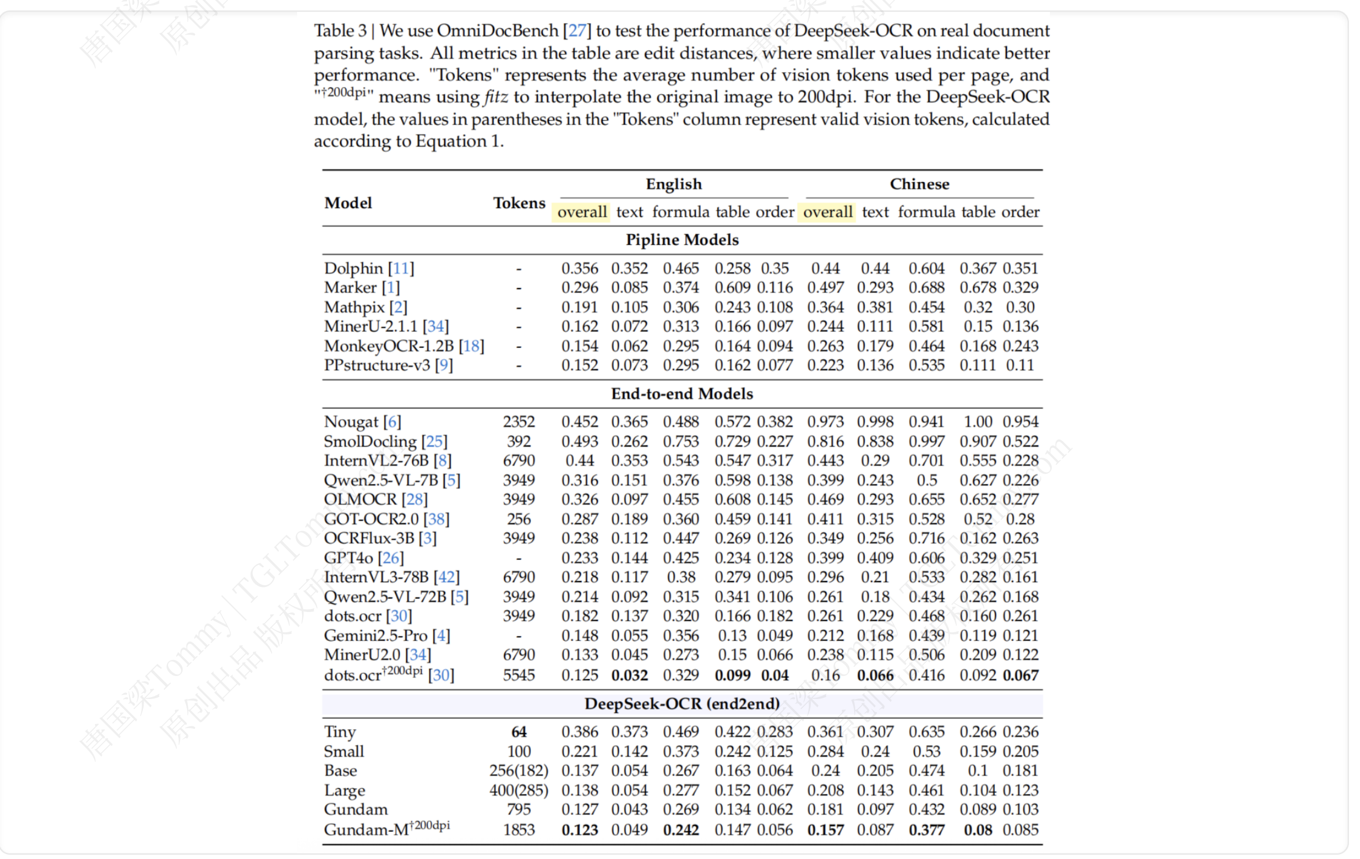


Table 3 | We use OmniDocBench [27] to test the performance of DeepSeek-OCR on real document parsing tasks. All metrics in the table are edit distances, where smaller values indicate better performance. "Tokens" represents the average number of vision tokens used per page, and "^{†200dpi}" means using *fitz* to interpolate the original image to 200dpi. For the DeepSeek-OCR model, the values in parentheses in the "Tokens" column represent valid vision tokens, calculated according to Equation 1.

Model	Tokens	English					Chinese				
		overall	text	formula	table	order	overall	text	formula	table	order
Pipline Models											
Dolphin [11]	-	0.356	0.352	0.465	0.258	0.35	0.44	0.44	0.604	0.367	0.351
Marker [1]	-	0.296	0.085	0.374	0.609	0.116	0.497	0.293	0.688	0.678	0.329
Mathpix [2]	-	0.191	0.105	0.306	0.243	0.108	0.364	0.381	0.454	0.32	0.30
MinerU-2.1.1 [34]	-	0.162	0.072	0.313	0.166	0.097	0.244	0.111	0.581	0.15	0.136
MonkeyOCR-1.2B [18]	-	0.154	0.062	0.295	0.164	0.094	0.263	0.179	0.464	0.168	0.243
PPstructure-v3 [9]	-	0.152	0.073	0.295	0.162	0.077	0.223	0.136	0.535	0.111	0.11
End-to-end Models											
Nougat [6]	2352	0.452	0.365	0.488	0.572	0.382	0.973	0.998	0.941	1.00	0.954
SmolDocling [25]	392	0.493	0.262	0.753	0.729	0.227	0.816	0.838	0.997	0.907	0.522
InternVL2-76B [8]	6790	0.44	0.353	0.543	0.547	0.317	0.443	0.29	0.701	0.555	0.228
Qwen2.5-VL-7B [5]	3949	0.316	0.151	0.376	0.598	0.138	0.399	0.243	0.5	0.627	0.226
OLMOCR [28]	3949	0.326	0.097	0.455	0.608	0.145	0.469	0.293	0.655	0.652	0.277
GOT-OCR2.0 [38]	256	0.287	0.189	0.360	0.459	0.141	0.411	0.315	0.528	0.52	0.28
OCRFlux-3B [3]	3949	0.238	0.112	0.447	0.269	0.126	0.349	0.256	0.716	0.162	0.263
GPT4o [26]	-	0.233	0.144	0.425	0.234	0.128	0.399	0.409	0.606	0.329	0.251
InternVL3-78B [42]	6790	0.218	0.117	0.38	0.279	0.095	0.296	0.21	0.533	0.282	0.161
Qwen2.5-VL-72B [5]	3949	0.214	0.092	0.315	0.341	0.106	0.261	0.18	0.434	0.262	0.168
dots.ocr [30]	3949	0.182	0.137	0.320	0.166	0.182	0.261	0.229	0.468	0.160	0.261
Gemini2.5-Pro [4]	-	0.148	0.055	0.356	0.13	0.049	0.212	0.168	0.439	0.119	0.121
MinerU2.0 [34]	6790	0.133	0.045	0.273	0.15	0.066	0.238	0.115	0.506	0.209	0.122
dots.ocr^{†200dpi} [30]	5545	0.125	**0.032**	0.329	**0.099**	**0.04**	0.16	**0.066**	0.416	0.092	**0.067**
DeepSeek-OCR (end2end)											
Tiny	**64**	0.386	0.373	0.469	0.422	0.283	0.361	0.307	0.635	0.266	0.236
Small	100	0.221	0.142	0.373	0.242	0.125	0.284	0.24	0.53	0.159	0.205
Base	256(182)	0.137	0.054	0.267	0.163	0.064	0.24	0.205	0.474	0.1	0.181
Large	400(285)	0.138	0.054	0.277	0.152	0.067	0.208	0.143	0.461	0.104	0.123
Gundam	795	0.127	0.043	0.269	0.134	0.062	0.181	0.097	0.432	0.089	0.103
Gundam-M^{†200dpi}	1853	**0.123**	0.049	**0.242**	0.147	0.056	**0.157**	0.087	**0.377**	**0.08**	0.085

性能亮点

- **Small 模式**：仅使用 100 个视觉词元，就超越了使用 256 个词元的 GOT-OCR2.0
- **Gundam 模式**：使用少于 800 个视觉词元，性能优于平均需要 6000+ 词元的 MinerU2.0
- **Gundam-M 模式**：在英文文档上达到 0.123 的编辑距离，处于端到端模型领先水平
- 展示了**巨大的实用价值**：单个 A100-40G GPU 每天可为 LLMs/VLMs 生成超过 20 万页训练数据

深度解析能力

结构化提取与理解

模型能够对文档内的图表、化学式、几何图形等进行结构化提取

图 7: 图表解析示例

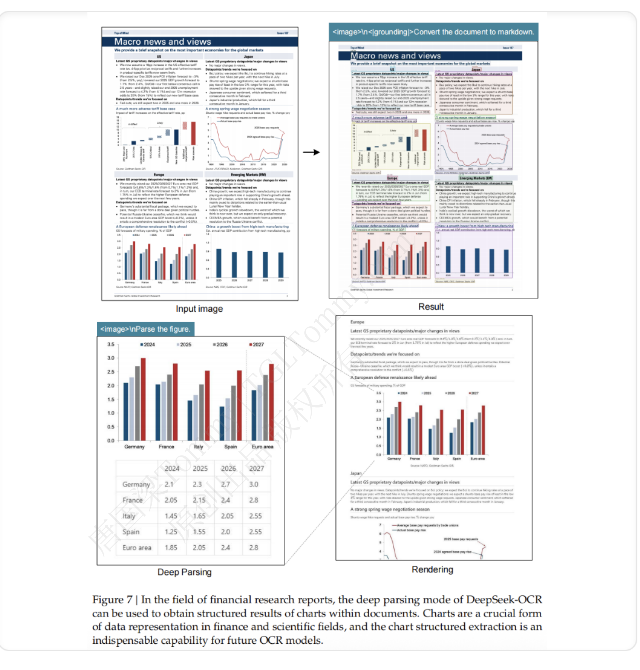


Figure 7 | In the field of financial research reports, the deep parsing mode of DeepSeek-OCR can be used to obtain structured results of charts within documents. Charts are a crucial form of data representation in finance and scientific fields, and the chart structured extraction is an indispensable capability for future OCR models.

将图表数据转换为 HTML 表格格式

化学分子式识别

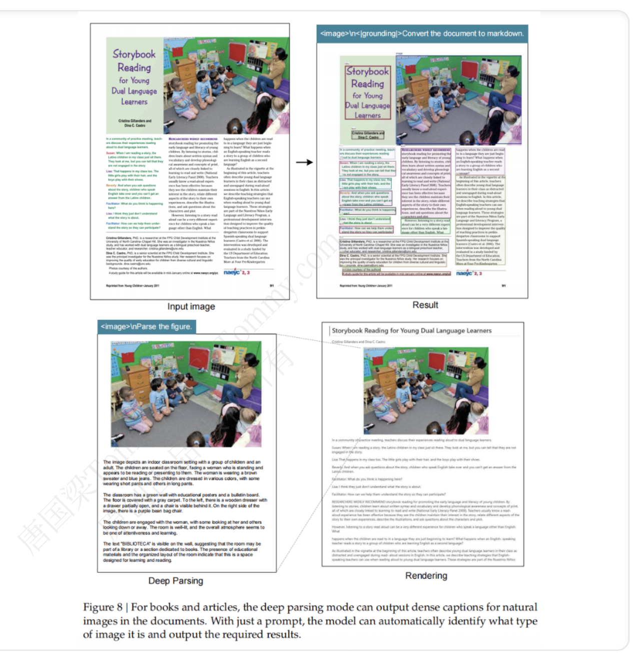


Figure 8 | For books and articles, the deep parsing mode can output dense captions for natural images in the documents. With just a prompt, the model can automatically identify what type of image it is and output the required results.

识别分子结构并转换为 SMILES 格式

实验结论

1. 可行性验证：成功验证了上下文光学压缩的可行性。模型能够从远少于原文词元数量的视觉词元中有效解码文本，在超过 10 倍的压缩率下仍保持高精度。

2. 实用价值：DeepSeek-OCR 是一个高度实用的模型，不仅在标准 OCR 基准测试上以更少的词元实现了 SOTA 性能，而且能够用于大规模的预训练数据生产。

3. 未来方向：光学压缩为解决 LLMs 中的长上下文挑战提供了一个充满潜力的新方向，有望通过协同结合视觉和语言模态来提升计算效率。