

# PR #7267 完整报告

PaddlePaddle/FastDeploy

[Docs] Update docs for release/2.5

合并时间: 2026-04-09 16:07

原文链接: <http://prhub.com.cn/PaddlePaddle/FastDeploy/pull/7267>

## 执行摘要

本 PR 同步了 FastDeploy v2.5 版本的中英文文档，更新了 README 发布信息和 NVIDIA GPU 安装指南，修复了残留的旧版本引用并澄清安装命令细节。作为纯文档变更，风险较低，但确保了用户能准确获取部署信息，维护了文档一致性。

## 功能与动机

为什么做: FastDeploy v2.5 发布后，原始提交仅更新了中文文档，导致英文文档仍引用 v2.4 版本，且安装指南中存在错误版本号（如 `paddlepaddle-gpu==3.3.0`）。作者 Jiang-Jia-Jun 在评论中明确指出：“这个 PR 仅修改了中文文档，请帮忙修改英文文档”，目标是消除版本不一致，提供准确的用户安装指引。

## 实现拆解

改动按文件拆解如下： | 文件 | 关键变更 | 影响 | |-----|-----|-----| | `README_CN.md` 和 `README_EN.md` | 新增 v2.5 发布条目（支持 Qwen3-VL MoE、W4AF8 量化等），调整 v2.4 条目格式 | 用户获取最新发布信息 | | `docs/get_started/installation/nvidia_gpu.md`（英文） | 更新 Docker 镜像至 2.5.0，支持 SM80/86/89/90；更新 `paddlepaddle-gpu` 至 3.3.1，增加 CUDA 12.9 选项；更新 `fastdeploy-gpu` 至 2.5.0，简化安装命令；添加说明澄清 `--extra-index-url` 仅用于依赖下载 | 英文用户安装指南 | | `docs/zh/get_started/installation/nvidia_gpu.md`（中文） | 同步以上变更，修复残留版本引用 | 中文用户安装指南 |

关键代码逻辑示例（从 patch 中提取）：`**Notice**`: The pre-built image supports SM80/86/89/90 GPUs (e.g. A800/H800/L20/L40/4090).

## 评论区精华

review 讨论中，最有价值的交锋围绕文档一致性和清晰度：

- 中英文不一致: `fastdeploy-bot` 指出：“英文版已将 Docker 镜像的 Notice 更新为支持 SM 80/86/89/90 架构，但中文版第 17 行的注意说明仍保留旧内容”，这会导致用户误操作。作者及时响应，Copilot 在提交中同步更新。
- 安装命令澄清: Copilot 评论：“文案强调‘不要通过 pypi 源安装’，但命令里又使用了 `--extra-index-url...` 这会让读者误解”，作者 Jiang-Jia-Jun 回复：“这个说得很对，补充说明下 `extra-index-url` 只用下依赖下载吧”。随后提交添加了明确说明，提升了文档可读性。

## 风险与影响

### 技术风险:

1. 文档不一致: 若中英文文档未同步, 用户可能基于错误信息部署 (如误判 GPU 支持范围), 但 review 过程已捕获并修复。
2. 误导风险: 安装命令中 `--extra-index-url` 若不澄清, 用户可能误从 PyPI 安装 `fastdeploy-gpu`, 导致版本冲突; 提交中的说明已缓解此风险。

### 影响分析:

- 用户: 所有 FastDeploy 用户受益于准确的 v2.5 文档, 减少安装错误。
- 系统: 无运行时影响, 纯文档变更。
- 团队: 维护文档一致性, 降低后续支持开销。

## 关联脉络

与近期 PR 7278 (添加 DeepSeek-V3.2 文档) 类似, 本 PR 是 FastDeploy 文档维护 workflow 的一部分, 反映团队在版本发布后及时同步文档的实践。从历史 PR 看, 文档更新常伴随发布周期 (如 v2.5), 旨在确保用户指南与代码演进同步。