

字幕TXT转SRT工具_使用说明书

一、工具介绍

本工具用于把音频或视频生成 .srt 字幕文件。你可以根据手头素材选择不同方式生成字幕：

- **离线识别**：软件打开后默认选择此模式。适合没有 TXT 文案、希望直接识别音频内容生成字幕的情况。
- **按停顿匹配TXT**：适合已经有 TXT 文案，只是缺少字幕时间轴的情况。它会根据音频里的说话停顿，把 TXT 内容按顺序匹配到时间段里。
- **豆包云识别**：适合已经配置豆包语音识别接口，希望通过云端识别生成字幕的情况。

默认打开后：

- 生成模式为 **离线识别**。
- **不勾选**“用 TXT 覆盖离线识别文本”。
- 也就是说，默认会直接使用离线识别出的文字生成字幕，不强制要求选择 TXT。

二、准备工作

1. 准备音频或视频文件

支持常见音频、视频文件，例如：

- 视频：MP4、MOV、MKV、AVI 等
- 音频：MP3、WAV、M4A 等

实际能否处理，取决于当前电脑里的 FFmpeg 是否支持该格式。

2. 准备 FFmpeg

工具需要 FFmpeg 来读取音视频。推荐把 FFmpeg 放在工具同级目录，例如：

```
工具目录/ffmpeg/ffmpeg.exe
```

或：

```
工具目录/ffmpeg/bin/ffmpeg.exe
```

如果电脑系统 PATH 里已经安装 FFmpeg，也可以正常使用。

3. 准备离线识别模型（仅离线识别需要）

使用“离线识别”时，需要本地可用的 faster-whisper 模型。常见放置方式：

```
工具目录/models/faster-whisper-模型名/
```

或：

```
工具目录/_fw_models/模型目录/
```

如果软件界面显示“未检测到本地模型”，需要先放入模型文件，或切换到“按停顿匹配TXT”模式使用已有 TXT 生成字幕。

三、界面区域说明

1. 文件列表区

用于添加要处理的音频或视频文件。

常用按钮：

- **添加文件**：选择一个或多个音频/视频文件。
- **添加文件夹**：选择整个文件夹，软件会自动加入里面支持的音视频文件。
- **移除选中**：从列表中移除当前选中的文件。
- **清空列表**：清空所有待处理文件。

如果当前环境支持拖拽，也可以把文件或文件夹直接拖入文件列表区。

2. 输出目录

用于设置生成的 `.srt` 字幕保存位置。

默认情况下，添加素材后软件会自动生成输出目录。你也可以手动点击 **选择...** 指定保存位置。

处理完成后，点击 **打开输出目录** 可以快速查看生成结果。

3. 生成模式

这里决定字幕生成方式。

离线识别

默认模式。软件会识别音视频里的语音内容，并生成字幕时间轴和字幕文本。

默认不勾选“用 TXT 覆盖离线识别文本”，因此不需要选择 TXT 文件。

适合：

- 没有字幕文案；
- 想直接从视频或音频中识别出字幕；
- 希望尽量减少手动准备文本。

按停顿匹配TXT

这个模式不会识别语音内容，只会分析音频中的说话停顿，再把 TXT 文本按顺序匹配到这些时间段里。

适合：

- 已经有完整 TXT 文案；
- 只是缺少 SRT 时间轴；
- 想快速生成字幕初稿。

注意：这个模式不理解文字内容。如果 TXT 断句和实际说话节奏差很多，可能出现字幕错位，需要调整 TXT 换行或后期微调。

豆包云识别

调用豆包语音识别接口生成字幕。需要填写 Access Token、APP ID、服务版本和资源 ID。

适合：

- 已经有豆包语音识别接口；
- 不想使用本地模型；
- 希望通过云端识别生成字幕。

四、TXT 设置说明

TXT 设置主要用于两种情况：

1. 选择 **按停顿匹配TXT** 模式；
2. 在识别模式中勾选 **用 TXT 覆盖识别文本**。

1. 指定一个 TXT

所有待处理文件共用同一个 TXT 文案。

适合单个视频，或多个视频确实共用同一份文案的情况。

2. 使用同名 TXT

每个媒体文件自动寻找同目录下同名的 `.txt` 文件。

例如：

```
课程01.mp4
课程01.txt
课程02.mp4
课程02.txt
```

选择“使用同名 TXT”后，软件会自动匹配对应文本。

五、推荐操作流程

场景一：没有 TXT，直接识别生成字幕

1. 打开软件。
2. 保持模式为 **离线识别**。
3. 不勾选 **用 TXT 覆盖离线识别文本**。
4. 点击 **添加文件** 或 **添加文件夹**。
5. 确认输出目录。
6. 点击 **开始生成**。
7. 完成后点击 **打开输出目录** 查看 `.srt` 文件。

场景二：已有 TXT，按停顿快速生成字幕

1. 添加音频或视频文件。
2. 生成模式选择 **按停顿匹配TXT**。
3. 在“字幕 TXT”区域选择 TXT：
 - 单个视频：选择 **指定一个 TXT**；
 - 批量视频：选择 **使用同名 TXT**。
4. 参数先保持默认。
5. 点击 **开始生成**。
6. 完成后打开输出目录查看结果。

场景三：用离线识别生成时间轴，但字幕文字用自己的 TXT

1. 添加音频或视频文件。
2. 选择 **离线识别**。
3. 勾选 **用 TXT 覆盖离线识别文本**。
4. 选择 TXT 文件，或选择使用同名 TXT。
5. 点击 **开始生成**。

这种方式适合：识别时间轴比较准，但你希望字幕文字使用自己整理过的文案。

场景四：使用豆包云识别生成字幕

- 1. 选择 豆包云识别。
- 2. 填写 Access Token。
- 3. 填写 APP ID。
- 4. 选择服务版本和资源 ID。
- 5. 根据需要决定是否勾选 用 TXT 覆盖云识别文本。
- 6. 添加文件并开始生成。

六、常用参数说明

按停顿匹配TXT相关参数

参数	默认值	说明
帧长(ms)	30	音频分析的小单位，一般不需要改。
静音判定(ms)	300	停顿多久算作可能的分段。数值越大，切得越少。
最短语音段(ms)	400	太短的声音段会合并，避免字幕过碎。
首尾扩展(s)	0.05	每条字幕前后稍微多留一点时间。

离线识别相关参数

参数	建议	说明
识别模型	使用已检测到的本地模型	模型越大通常越准，但速度更慢。
设备	CPU	没有显卡环境时用 CPU 更稳。
精度	int8	CPU 下速度和占用更友好。
仅用本地模型	勾选	避免运行时联网下载模型。
语言	zh	中文素材建议填写 zh。
过滤长静音	默认不勾选	特殊长静音素材可尝试开启。
最短子段(ms)	260	避免生成过短字幕。

豆包云识别相关参数

参数	说明
Access Token	豆包接口访问凭证。
APP ID	豆包应用 ID。

参数	说明
服务版本	极速版或标准版。
资源 ID	当前服务对应的资源 ID。
语言	中文通常填写 zh。

七、结果文件说明

处理完成后，每个媒体文件会生成一个同名 `.srt` 文件。

例如：

```
输入：课程01.mp4
输出：课程01.srt
```

生成的 SRT 可以导入剪映、PR、播放器或其他支持 SRT 字幕的软件。

八、常见问题

1. 为什么打开后不需要选 TXT？

因为软件默认进入 **离线识别**，并且默认不勾选“用 TXT 覆盖离线识别文本”。这种情况下，软件会直接识别音视频内容生成字幕，所以 TXT 不是必需项。

2. “按停顿匹配TXT”和“离线识别”有什么区别？

按停顿匹配TXT：不识别内容，只根据声音停顿给 TXT 配时间轴。

离线识别：会识别音频里说了什么，并生成字幕内容和时间轴。

有 TXT 文案时，可以用“按停顿匹配TXT”；没有 TXT 时，用“离线识别”。

3. 按停顿匹配TXT生成的字幕错位怎么办？

可以尝试：

- 调整 TXT 换行，让每行更接近实际一句话；
- 把过长的一行拆成多行；
- 把过短、过碎的行合并；
- 适当调整“静音判定(ms)”参数。

4. 提示未检测到本地模型怎么办？

说明当前没有找到可用的 faster-whisper 模型。可以：

- 把模型目录放到工具同级的 `models` 或 `_fw_models` 文件夹；
- 确认模型目录里包含必要模型文件；
- 暂时切换到“按停顿匹配TXT”模式处理已有 TXT 文案。

5. 提示找不到 FFmpeg 怎么办？

请把 `ffmpeg.exe` 放到以下任一位置：

```
工具目录/ffmpeg/ffmpeg.exe  
工具目录/ffmpeg/bin/ffmpeg.exe
```

或安装 FFmpeg 并加入系统 PATH。

6. 批量处理时怎么配 TXT？

如果每个视频都有自己的 TXT，建议使用“使用同名 TXT”。确保文件名一致，例如：

```
视频A.mp4  
视频A.txt
```

文件名不一致时，软件会提示找不到同名 TXT。

九、建议使用方式

新用户建议按这个顺序使用：

1. 没有 TXT：直接用默认 **离线识别**。
2. 已有 TXT：切换到 **按停顿匹配TXT**。
3. 有 TXT 且想要更智能时间轴：用 **离线识别 + 勾选 TXT 覆盖**。
4. 有豆包接口：再考虑 **豆包云识别**。