

[Python 原创] 【Python】一键爬取快手所有关注账号的全部视频 [复制链接]

FirmTao 2021-12-23 11:53

楼主 电梯直达

本帖最后由 FirmTao 于 2021-12-25 12:00 编辑

PS：更新了一下，之前防盗链忘记改了，现在需要添加自己的

最开始只是为了下载快手关注的一个账号的全部视频，写完代码后又想下载全部关注的视频😓，
但又不想一个一个的手动操作，所有一想直接一键搞定，走起

以下为抓包过程：

先进入网站，网址 快手官网，不知道的自己去百度

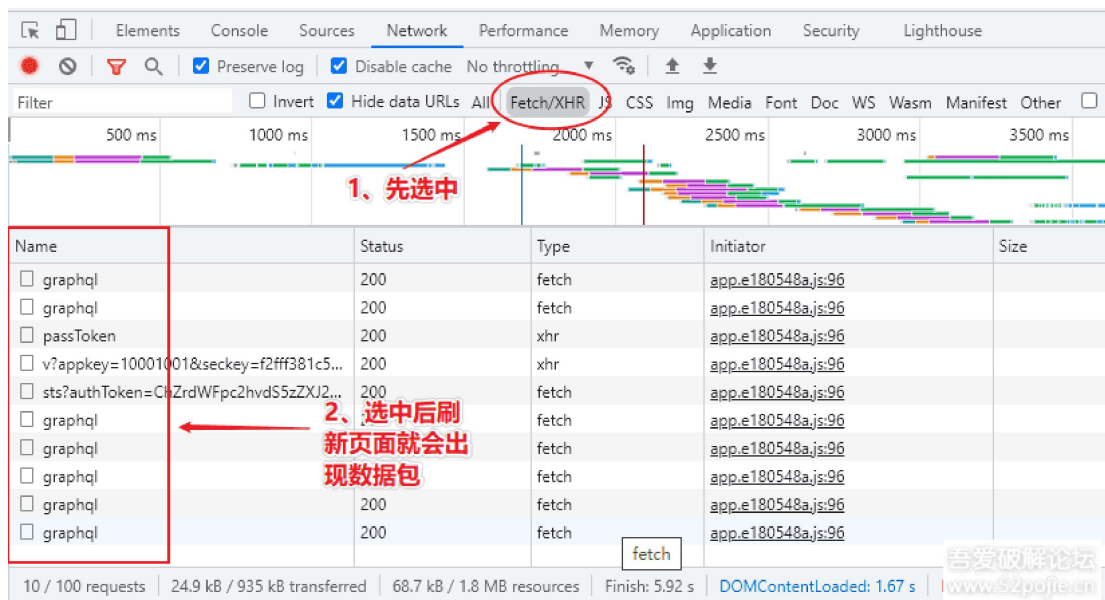
1、进入后登录自己的账号，然后点击进入主页



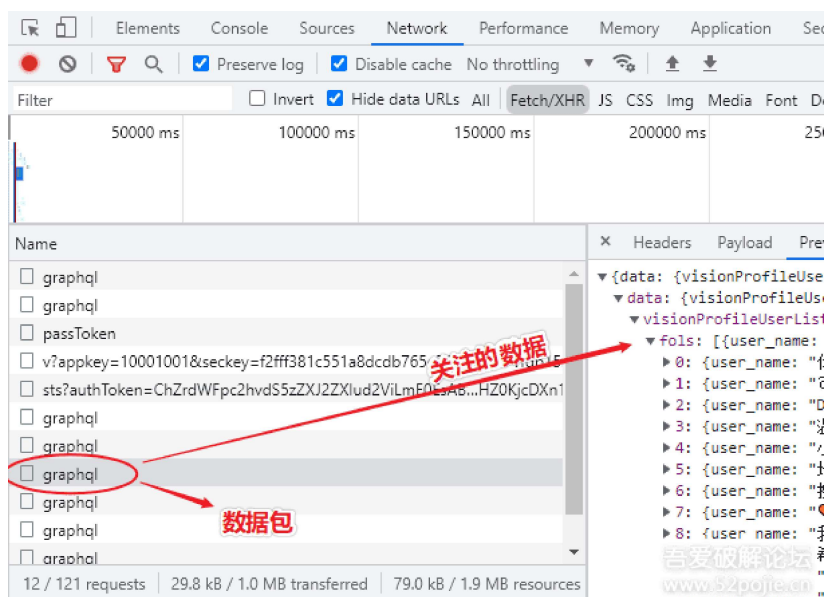
进入后左侧可以看到自己的关注人数，右侧是具体的关注账号，我们要获取的就是右侧的数据



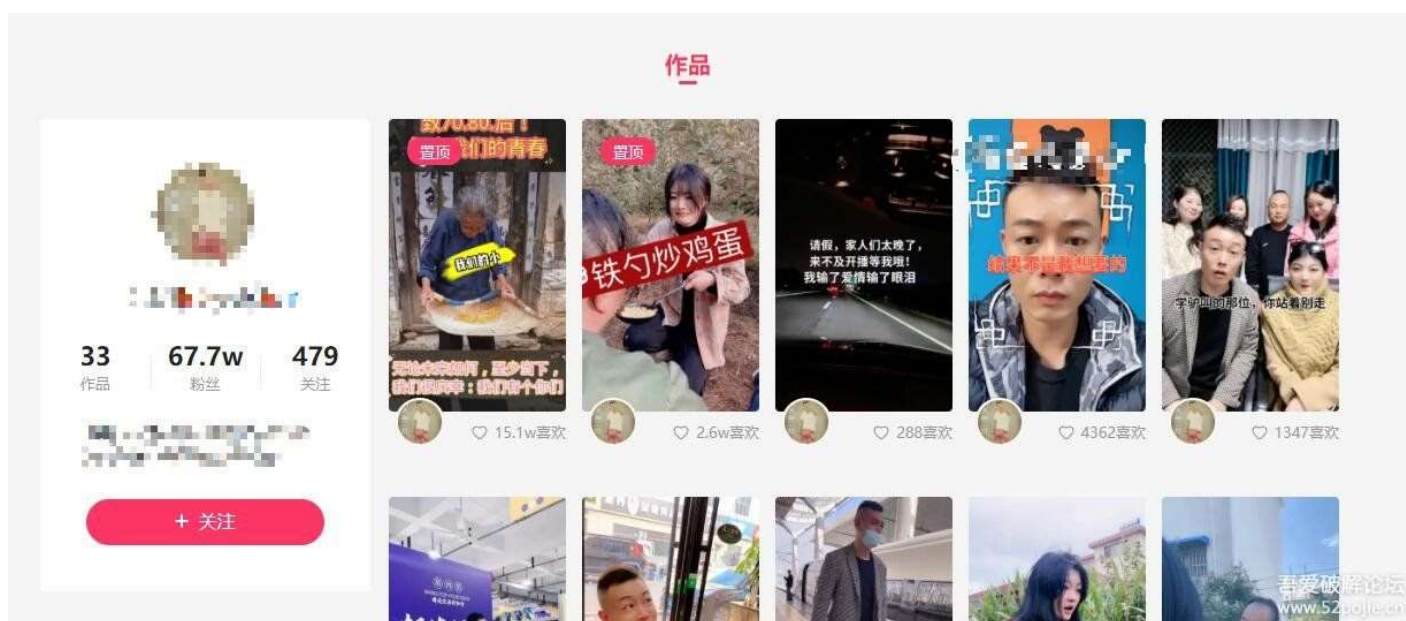
2、右键打开检测，切换到 network 窗口，再选中第三行工具栏的 Fetch/XHR，然后刷新一下当前页面，可以看到数据包



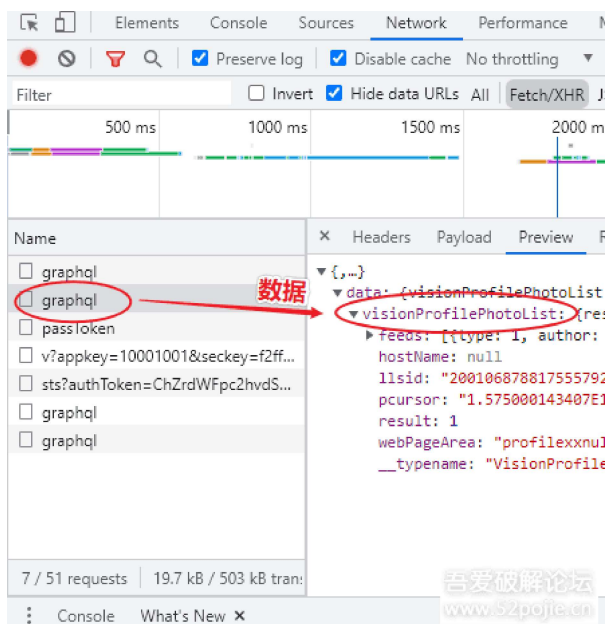
3、依次点击 graphql 数据包，找到需要的数据，这个列表就是我们需要的关注数据 `visionProfileUserList`



4、这个关注的用户获取到后，之后就是去找个人所有视频的数据啦啦啦，随便打开一个人的主页，可以看到全部视频



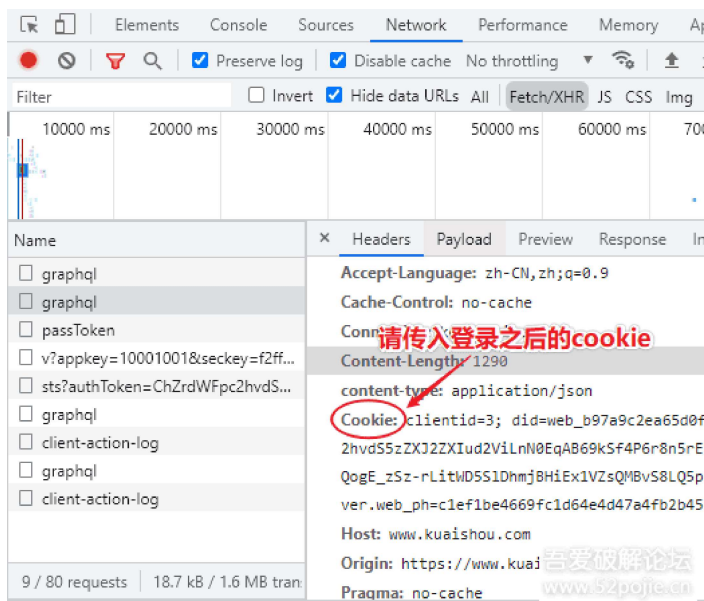
5、跟第 2 步一样，刷新出数据包，然后依次点击 graphql 数据包，找到需要的数据



6、OK，数据全部找到，接下来就是写代码了，话不多说直接上代码

代码可能写的不好，本人刚接触，大佬请指正批评，让我进步

提示：有四个参数，你只需要填两个 一个是Cookie Cookie填的是登录后的，另一个是User-Agent



User-Agent 在Cookie的下面!!!

7、这是运行效果图

```
python.exe "E:/PythonProject/04快手爬虫/kuaishou.py"

第1页:no_more
1、 拾芥: 3x67n3u5aik32w4 >>> ID获取成功!!!
2、 小荔枝妹破50w: 3xcqxz62ga5hfnw >>> ID获取成功!!!
3、 夏枳美女合集: 3xu7brmsbyehqh6 >>> ID获取成功!!!
['3x67n3u5aik32w4', '3xcqxz62ga5hfnw', '3xu7brmsbyehqh6']
第1位关注: 3x67n3u5aik32w4 全部视频下载中...
1.637501098848E12
1、 我真的很想你声控拾芥 >>> 下载完成!!!
2、 还是算了吧我没有那个好运气拾芥声控 >>> 下载完成!!!
3、 不如我们从头来过声控拾芥 >>> 下载完成!!!
4、 我要送你满屋热气与一把蝴蝶声控拾芥 >>> 下载完成!!!
5、 我们意念合一声控拾芥 >>> 下载完成!!!
6、 江南的早梅还开着只是你再不折下那枝梅我就要走了声控拾芥 >>> 下载完成!!!
7、 太可笑了是不是声控拾芥 >>> 下载完成!!!
8、 似是故人来王春荣耀配音拾芥 >>> 下载完成!!!
9、 从此天光大亮你是我全部的渴望与幻想声控拾芥 >>> 下载完成!!!
10、 我们终会再次相遇对吧声控拾芥 >>> 下载完成!!!
11、 我想我会永远爱他声控拾芥 >>> 下载完成!!!
12、 我今天讨厌所有人但是不包括你声控拾芥 >>> 下载完成!!!
13、 那便珍重吧声控拾芥 >>> 下载完成!!!
14、 时至今日我仍觉得分享欲是最高级别的浪漫声控拾芥 >>> 下载完成!!!
15、 如果你有空的话能不能陪我过个冬声控拾芥 >>> 下载完成!!!
```

吾爱破解论坛
www.52pojie.cn



8、最后请各位兄弟姐妹们有什么给什么，帮忙顶一顶，出问题可以评论区留言。多说一句，python是3.9版本

[Asm] 纯文本查看 复制代码

```
001 # coding:utf-8
002 import json
003 import os
004 import pprint
005 import re
006 import string
007 import sys
008 import time
009 import ctypes
010 import os
011 import platform
012 from zhon.hanzi import punctuation
013 import requests
014 import urllib3
015
016 urllib3.disable_warnings()
017
```

?

```

018
019 # 请求网页
020 def req_data(url, id, pcursor, ck, ua):
021     # 请求头
022     headers = {
023         'content-type': 'application/json',
024         'Cookie': ck,
025         'Host': 'www.kuaishou.com',
026         'Origin': 'https://www.kuaishou.com',
027         'Referer': 'https://www.kuaishou.com/profile/' + id,
028         'User-Agent': ua
029     }
030     # 请求参数
031     data = {
032         'operationName': 'visionProfilePhotoList',
033         'query': "query visionProfilePhotoList($pcursor: String, $userId: String, $page: String, $webPageArea: "
034             "String) {\n  visionProfilePhotoList(pcursor: $pcursor, userId: $userId, page: $page, webPageArea: "
035             "$webPageArea) {\n    result\n      llsid\n      webPageArea\n      feeds {\n        type\n        author {\n "
036             "      id\n      name\n      following\n      headerUrl\n      headerUrls {\n        cdn\n "
037             "      url\n      __typename\n      }\n      __typename\n      }\n      tags {\n        "
038             "type\n      name\n      __typename\n      }\n      photo {\n        id\n        duration\n        "
039             "caption\n        likeCount\n        realLikeCount\n        coverUrl\n        coverUrls {\n        "
040             "      cdn\n      url\n      __typename\n      }\n      photoUrls {\n        cdn\n        "
041             "url\n      __typename\n      }\n      photoUrl\n      liked\n      timestamp\n        "
042             "expTag\n      animatedCoverUrl\n      stereoType\n      videoRatio\n        "
043             "profileUserTopPhoto\n      __typename\n      }\n      canAddComment\n      currentPcursor\n        "
044             "llsid\n      status\n      __typename\n      }\n      hostName\n      pcursor\n      __typename\n      }\n}\n",
045         'variables': {'userId': id, 'pcursor': pcursor, 'page': 'profile'}
046     }
047     data = json.dumps(data)
048     data_json = requests.post(url=url, headers=headers, data=data, timeout=6.05).json()
049     # pprint.pprint(data_json)
050     return data_json
051
052
053 # 清洗文件名
054 def rep_char(chars):
055     eg_punctuation = string.punctuation
056     ch_punctuation = punctuation
057     # print("所有标点符号: ", eg_punctuation, ch_punctuation)
058     for item1 in eg_punctuation:
059         chars = chars.replace(item1, '')
060     for item2 in ch_punctuation:
061         chars = chars.replace(item2, '')
062     chars = chars.replace(' ', '').replace('\n', '').replace('\xa0', '').replace('\r', '')
063     return chars
064
065
066 # 磁盘内存检查
067 def get_free_space():
068     folder = os.path.abspath(sys.path[0])
069     if platform.system() == 'Windows':
070         free_bytes = ctypes.c_ulonglong(0)
071         ctypes.windll.kernel32.GetDiskFreeSpaceExW(ctypes.c_wchar_p(folder), None, None, ctypes.pointer(free_bytes))
072         return free_bytes.value / 1024 / 1024 / 1024
073     else:
074         st = os.statvfs(folder)
075         return st.f_bavail * st.f_frsize / 1024 / 1024
076
077
078 # 保存数据
079 def save(url, page, ck, ua, selfid):
080     except_lit = []
081     count = 0
082     idlist = get_all_ids(url, page, ck, ua, selfid)
083     for id in idlist:
084         count = count + 1
085         print(f'第{count}位关注: {id} 全部视频下载中...')
086         num = 0
087         # 循环下载视频, 直到 page == 'no_more'

```



```

088 while page != 'no_more':
089     time.sleep(1)
090     data = req_data(url, id, page, ck, ua)
091     # 获取翻页的参数
092     next_page_Pcursor = data['data']['visionProfilePhotoList']['pcursor']
093     page = next_page_Pcursor
094     print(next_page_Pcursor)
095     data_list = data['data']['visionProfilePhotoList']['feeds']
096     for item in data_list:
097         num = num + 1
098         video_name = item['photo']['caption']
099         video_url = item['photo']['photoUrl']
100         author = item['author']['name']
101         author = rep_char(author)
102         video_name = rep_char(video_name)
103         path = './video1'
104         if not os.path.exists(path + '/' + author + '/'):
105             os.makedirs(path + '/' + author + '/')
106         filepath = path + '/' + author + '/' + str(num) + '.' + video_name + '.mp4'
107         if os.path.exists(filepath):
108             print(f'{num}、 {video_name} >>> 已存在!!!')
109             # time.sleep(1)
110             continue
111         # 请求二进制视频数据
112         try:
113             video_content = requests.get(url=video_url, timeout=(3, 7)).content
114         except:
115             strss = f'{author}_{num}video_name: {video_url}'
116             except_lit.append(strss)
117             continue
118         with open(filepath, mode='wb') as f:
119             f.write(video_content)
120         print(f'{num}、 {video_name} >>> 下载完成!!!')
121         # 判断剩余容量是否充足
122         free_space = get_free_space()
123         if free_space <= 1:
124             break
125         # pcursor = page 这个变量的值必须为空，不用动他，它是换页的参数
126         page = ''
127         print(f'第{count}位关注: {id} 全部视频下载完成!!!')
128     with open('yc_info.txt', 'a') as f:
129         f.write(except_lit)
130         print('异常保存成功')
131     print(except_lit)
132
133
134 # 获取全部关注页面数据
135 def req_follow_data(url, pcursor, ck, ua, selfid):
136     # 请求头
137     headers = {
138         'content-type': 'application/json',
139         'Cookie': ck,
140         'Host': 'www.kuaishou.com',
141         'Origin': 'https://www.kuaishou.com',
142         'Referer': 'https://www.kuaishou.com/profile/' + selfid,
143         'User-Agent': ua
144     }
145     # 请求参数
146     data = {
147         'operationName': 'visionProfileUserList',
148         'query': 'query visionProfileUserList($pcursor: String! $ftype: Int!) {\n  visionProfileUserList(pcursor: '\n\n    '$pcursor, ftype: $ftype) {\n    result\n    folds {\n      user_name\n      headurl\n      '\n\n    'user_text\n      isFollowing\n      user_id\n      __typename\n    }\n    hostName\n    pcursor\n    '\n\n    ' __typename\n  }\n}\n',
149         'variables': {'ftype': 1, 'pcursor': pcursor}
150     }
151
152     data = json.dumps(data)
153     follow_json = requests.post(url=url, headers=headers, data=data).json()
154     # pprint.pprint(follow_json)
155     return follow_json

```

```

158
159
160 # 获取全部关注的id
161 def get_all_ids(url, page, ck, ua, selfid):
162     id_list = []
163     num = sign = 0
164     # 循环保存id, 直到 Pcursor == 'no_more'
165     while page != 'no_more':
166         time.sleep(1)
167         follow_data = req_follow_data(url, page, ck, ua, selfid)
168         # 获取翻页的参数
169         next_pcursor = follow_data['data']['visionProfileUserList']['pcursor']
170         page = next_pcursor
171         sign = sign + 1
172         print(f'第{sign}页:{next_pcursor}')
173         fols_list = follow_data['data']['visionProfileUserList']['fols']
174         for item in fols_list:
175             num = num + 1
176             user_name = item['user_name']
177             user_id = item['user_id']
178             id_list.append(user_id)
179             print(f'{num}、 {user_name}: {user_id} >>> ID获取成功!!!')
180     print(id_list)
181     print(id_list)
182     return id_list
183
184
185 if __name__ == '__main__':
186     link = 'https://www.kuaishou.com/graphql'
187     # pcursor这个变量的值开始必须为空, 不用动他, 它是翻页的参数
188     # selfid 是自己账号网址的最后面那一串 例如 https://www.kuaishou.com/profile/3xkfgnn9hkacbw selfid 就是 3xkfgnn9hkacbw
189     selfid = ''
190     pcursor = ''
191     # ck = '' 引号中间填登录后的 Cookie 值
192     ck = ''
193     # ua = '' 引号中间填 User-Agent
194     ua = ''
195     save(link, pcursor, ck, ua, selfid)

```

免费评分

	吾爱币	热心值	理由	收起
 A4碗		+ 1	欢迎分析讨论交流, 吾爱破解论坛有你更精彩!	
 春又来人已去		+ 1	热心回复!	
 h07799486	+ 1	+ 1	谢谢@Thanks!	
 m16139		+ 1	我很赞同!	
 随遇而安8	+ 1	+ 1	希望多多分享, 大家互相学习	
 三木零	+ 1	+ 1	用心讨论, 共获提升!	
 某些人	+ 1	+ 1	谢谢@Thanks!	
 onlycky	+ 1	+ 1	鼓励转贴优秀软件安全工具和文档!	
 Hiaappy	+ 1	+ 1	我很赞同!	
查看全部评分				

发帖前要善用【论坛搜索】功能，那里可能会有你要找的答案或者已经有人发布过相同内容了，请勿重复发帖。

	回复		举报
<div><div><div>破吾解爱</div></div></div>	<div><div>dmq336</div><div>2021-12-23 14:34</div></div> <div>感谢楼主分享。</div> <div><div>【吾爱破解论坛总版规】 - (让你充分了解吾爱破解论坛行为规则)</div></div> <div><div>回复</div><div>支持 1</div></div>		推荐
<div><div><div>破吾解爱</div></div></div>	<div><div>Hiaappy</div><div>2021-12-23 13:36</div></div> <div>吾爱破解论坛没有任何官方QQ群，禁止留联系方式，禁止任何商业交易。</div> <div>收藏了，很好的软件</div> <div>如何升级？如何获得积分？积分对应解释说明！</div> <div><div>回复</div><div>支持 1</div></div>		推荐
<div><div><div>破吾解爱</div></div></div>	<div><div>NNYY</div><div>2021-12-23 12:19</div></div> <div>《站点帮助文档》有什么问题来这里看看吧，这里有你想知道的内容！</div> <div>感谢楼主分享！</div> <div>呼吁大家发布原创作品添加吾爱破解论坛标识！</div> <div><div>回复</div><div>支持</div></div>		沙发
<div><div><div>破吾解爱</div></div></div>	<div><div>SDU123</div><div>2021-12-23 12:22</div></div> <div>感谢楼主分享</div> <div>如何快速判断一个文件是否为病毒！</div> <div><div>回复</div><div>支持</div></div>		3#
<div><div><div>lvcaolhx</div></div></div>	<div><div>lvcaolhx</div><div>2021-12-23 12:48</div></div> <div>楼主分享下，怎样模拟登陆拿到cookies?</div> <div><div>回复</div><div>支持</div></div>		4#
<div><div><div>破吾解爱</div></div></div>	<div><div>miaomiaofu</div><div>2021-12-23 13:16</div></div> <div>谢谢分享，</div> <div><div>回复</div><div>支持</div></div>		5#
<div><div><div>破吾解爱</div></div></div>	<div><div>rewin</div><div>2021-12-23 13:35</div></div> <div>麻烦问一下，下载下来的带快手水印吗？</div> <div><div>回复</div><div>支持</div></div>		6#
<div><div><div>破吾解爱</div></div></div>	<div><div>FirmTao</div><div>2021-12-23 14:08</div></div> <div><div><div>lvcaolhx</div>发表于 2021-12-23 12:48</div><div>楼主分享下，怎样模拟登陆拿到cookies?</div></div> <div>cookie是登录自己的账号后直接F12找到的啊，上面有图，至于你说的模拟登录要用到selenium，我这里不需要模拟登录啊</div> <div><div>回复</div><div>支持</div></div>		8#
<div><div><div>破吾解爱</div></div></div>	<div><div>FirmTao</div><div>2021-12-23 14:09</div></div> <div><div><div>rewin</div>发表于 2021-12-23 13:35</div><div>麻烦问一下，下载下来的带快手水印吗？</div></div> <div>没有水印哦，反正我下载没有水印</div> <div><div>回复</div><div>支持</div></div>		9#
<div><div><div>破吾解爱</div></div></div>	<div><div>ychiou</div><div>2021-12-23 14:23</div></div> <div>有成品吗？</div>		10#

高级模式

您需要登录后才可以回帖
 [登录](#) | [注册\(Register\)](#)


[用QQ帐号登录](#)

[发表回复](#)

警告：本版块禁止灌水或回复与主题无关内容，违者重罚！
☐ 回帖并转播
 ☐ 回帖后跳转到最后一页