

# 我把一个 77MB 的 PDF 电子书，变成了 23 MB，还更清晰了

作者: [HaujetZhao](#)

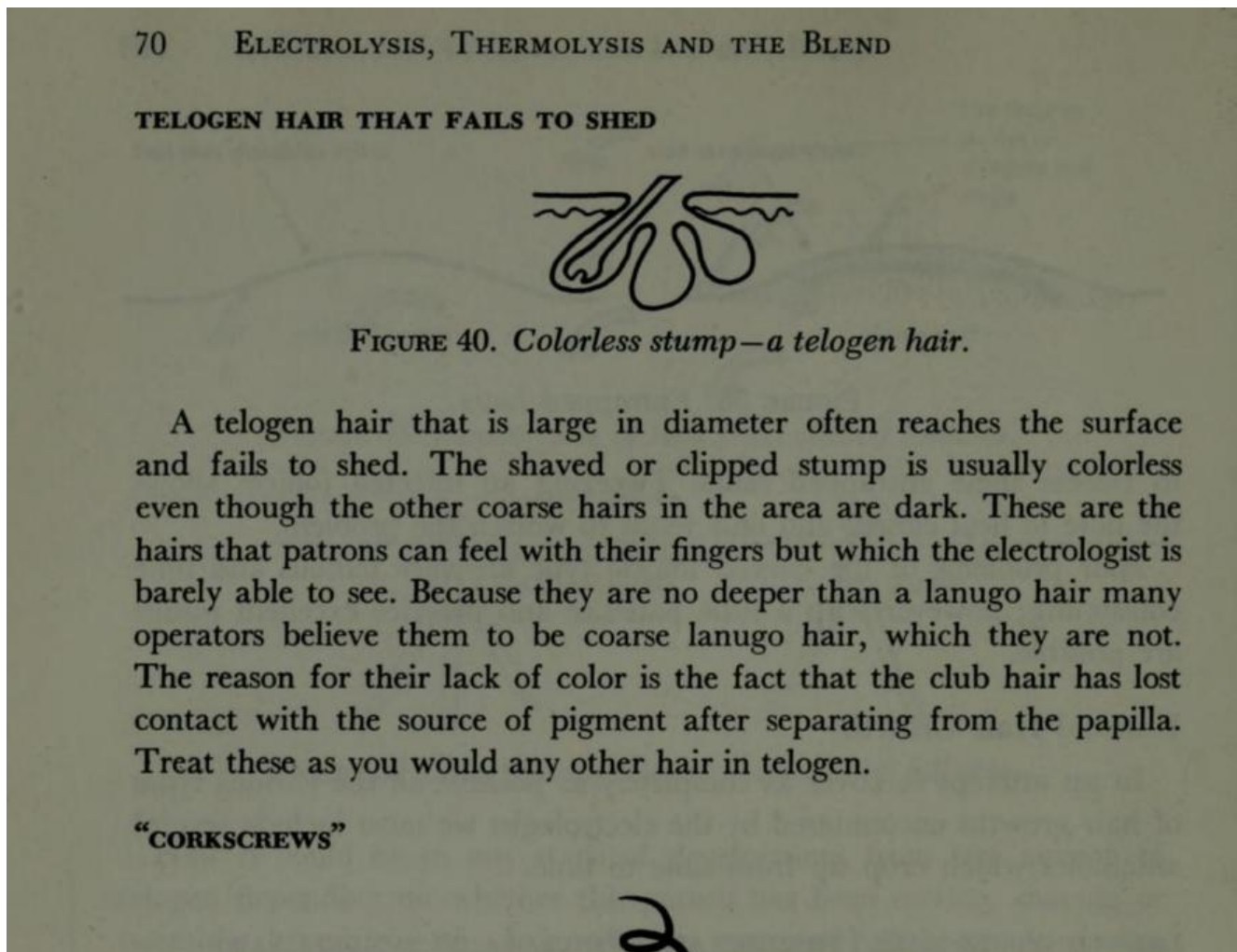
原文链接: <https://ld246.com/article/1625632016217>

来源网站: [链滴](#)

许可协议: [署名-相同方式共享 4.0 国际 \(CC BY-SA 4.0\)](#)

## 效果预览

这是原始 PDF:



这是优化后:

## TELOGEN HAIR THAT FAILS TO SHED



FIGURE 40. *Colorless stump—a telogen hair.*

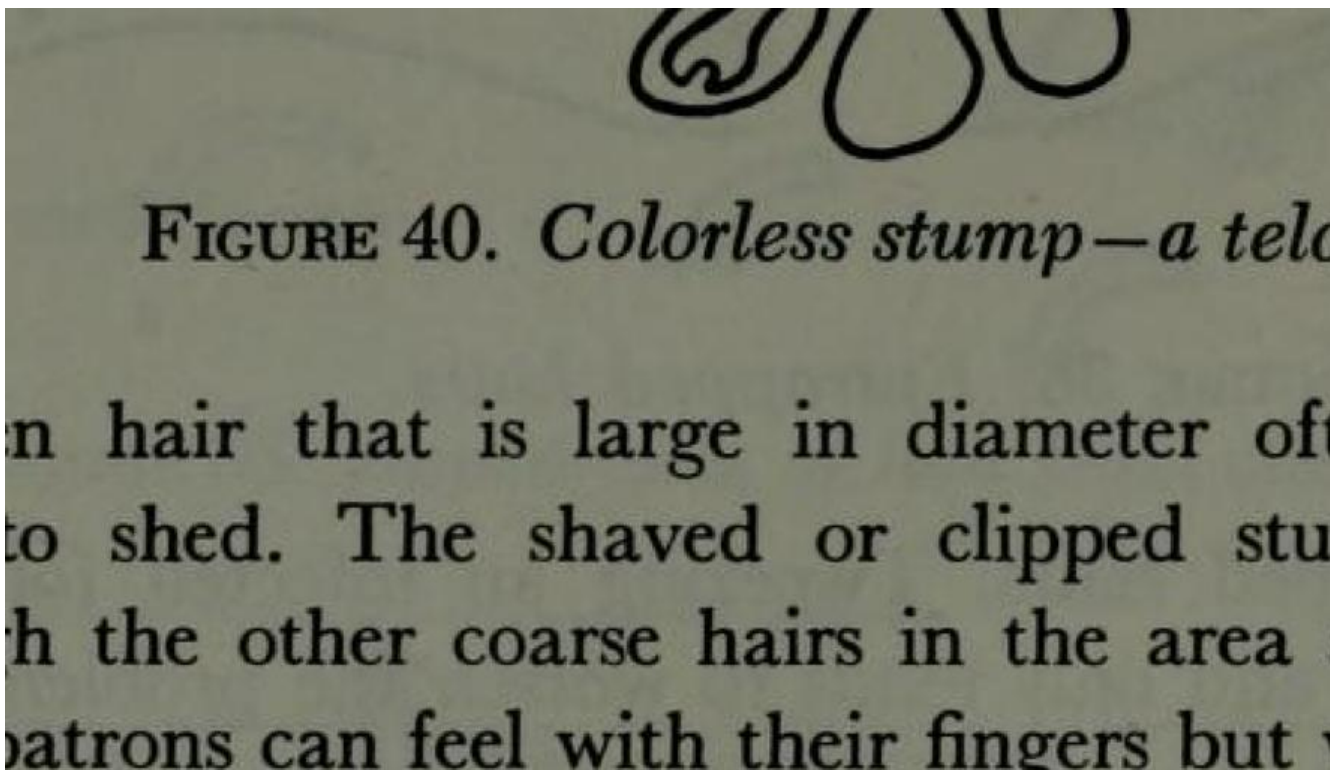
A telogen hair that is large in diameter often reaches the surface and fails to shed. The shaved or clipped stump is usually colorless even though the other coarse hairs in the area are dark. These are the hairs that patrons can feel with their fingers but which the electrologist is barely able to see. Because they are no deeper than a lanugo hair many operators believe them to be coarse lanugo hair, which they are not. The reason for their lack of color is the fact that the club hair has lost contact with the source of pigment after separating from the papilla. Treat these as you would any other hair in telogen.

### **“CORKSCREWS”**

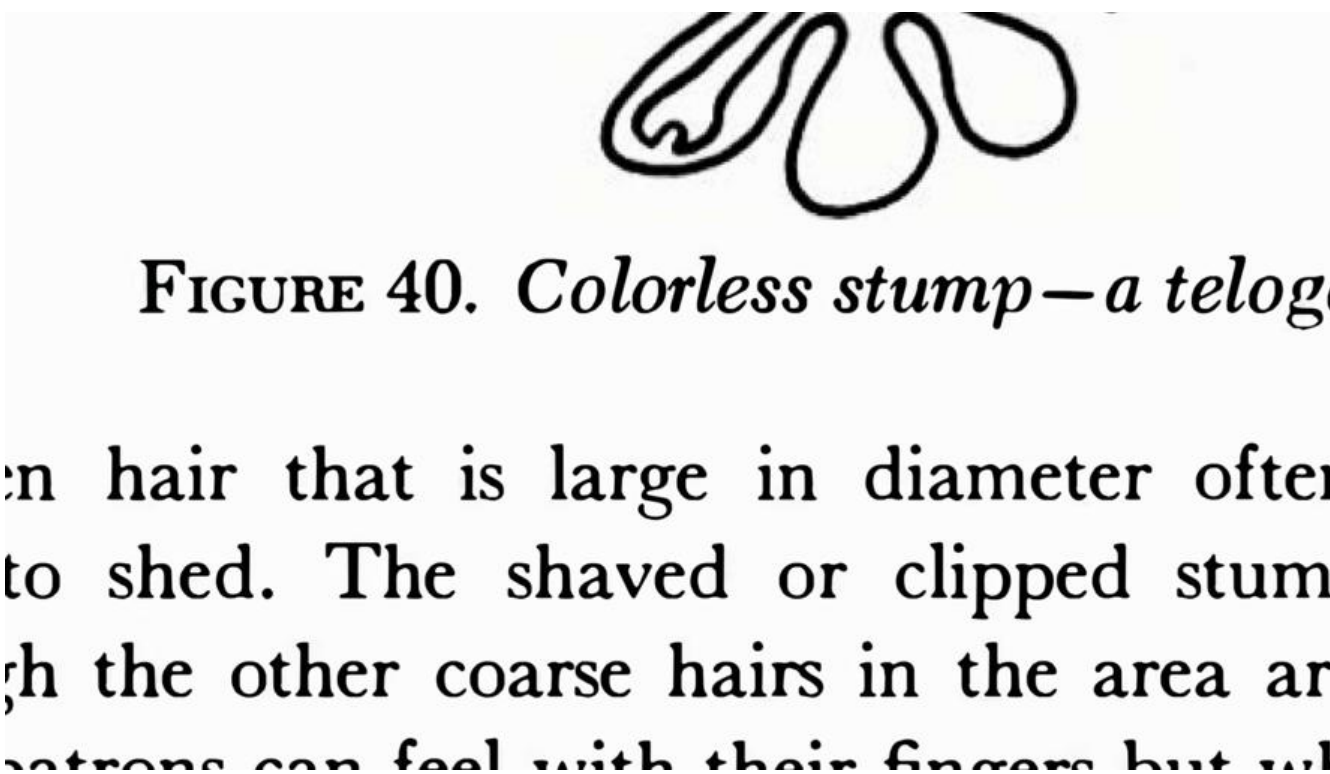


放大查看细节

原始 PDF 放大:



优化后放大:



### 为什么能缩小体积的同时还变清晰

原始 PDF 有 328 页，每一页都是裁切后的照片，分辨率为 1053 x 1573，DPI 72，即使平均每一页小为 250KB，那么整个 PDF 文件也有 82MB。

如果里面的所有文字，都是使用的矢量字体，那么体积就会小到几 MB 不到，就像是一个百万字的 txt 小说也不过 1MB 而已。

其实优化后的 PDF，就是将图片进行了优化，去除背景，变成纯白背景，降噪，合成为 PDF，再 OCR，将图片里面的文字部分替换为矢量文字。

## 优化流程

### 工具

- Adobe Acrobat
- Photoshop
- Waifu2x Extention GUI

### 提取照片

用 Acrobat 打开 PDF，导出为 JPG

### 照片优化

使用 Waifu2x Extention GUI，将图片进行 2x 放大，降噪 2 级，图片类型选择 2D 动漫，输出格式为 jpg，质量 100

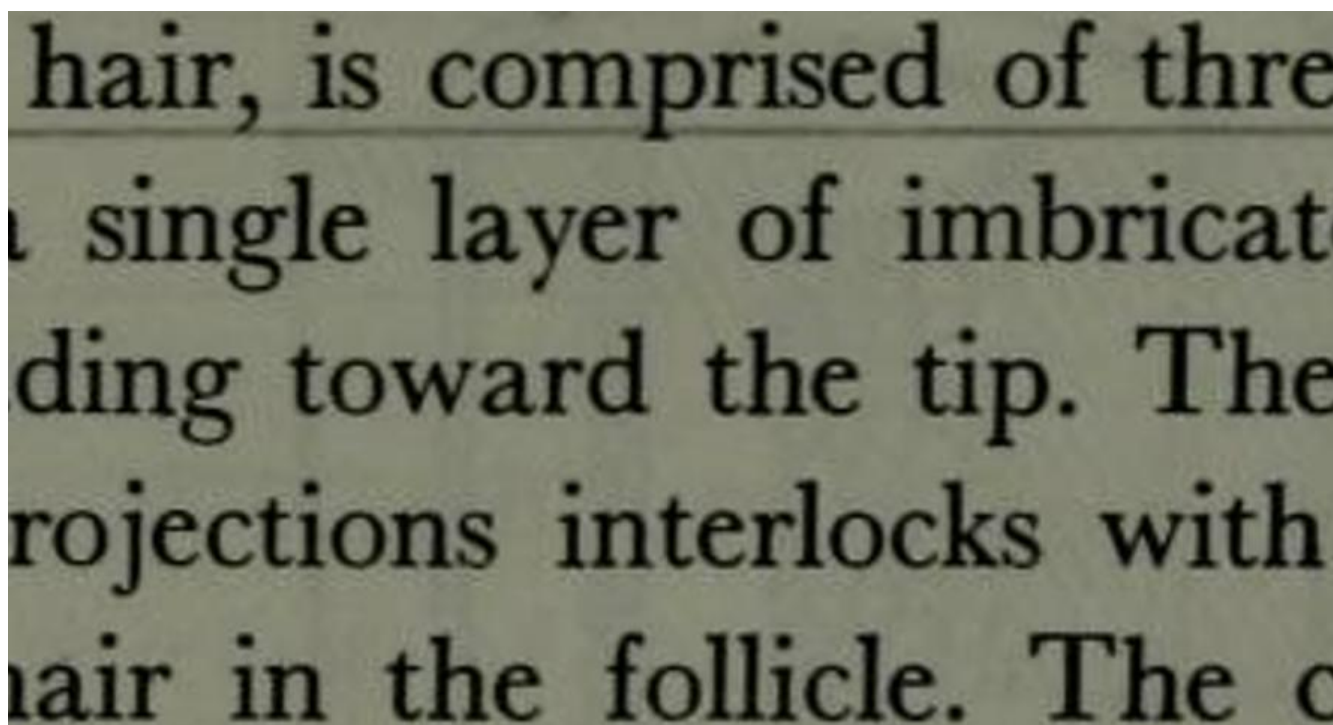
不能使用 PNG 输出，因为 PNG 在 Acrobat 中合并为 PDF 时，会被转换为 JPG，并且会降低品质导致字体周围有噪点，使得 OCR 的效果下降，生成的矢量字体曲线不够光滑，观感差。

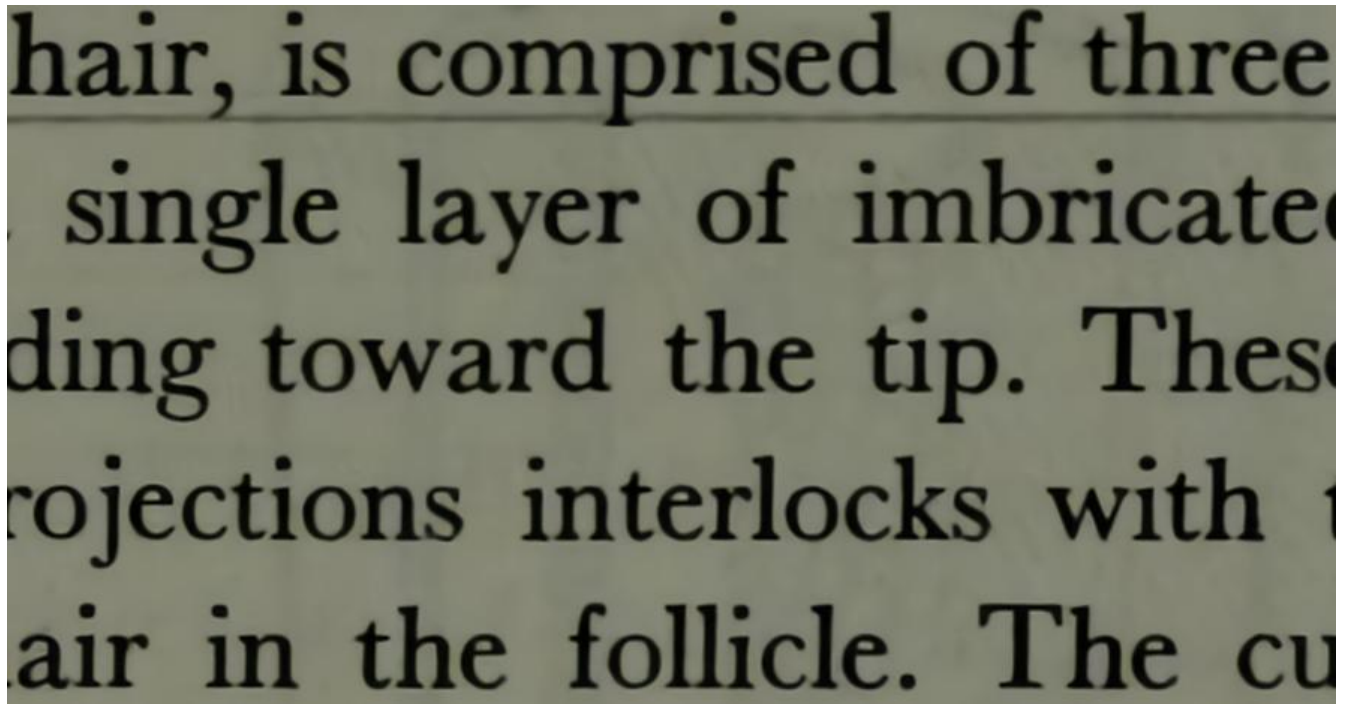
Waifu2x 是一个人工智能图片放大模型，可以生成效果非常好的放大图片

将图片放大，是为了让 OCR 生成的矢量字体边缘更平滑，看起来更舒服。如果图片分辨率太小、噪多，OCR 生成的矢量字体就会边缘粗糙、观感甚至不如原来低分辨率的图片

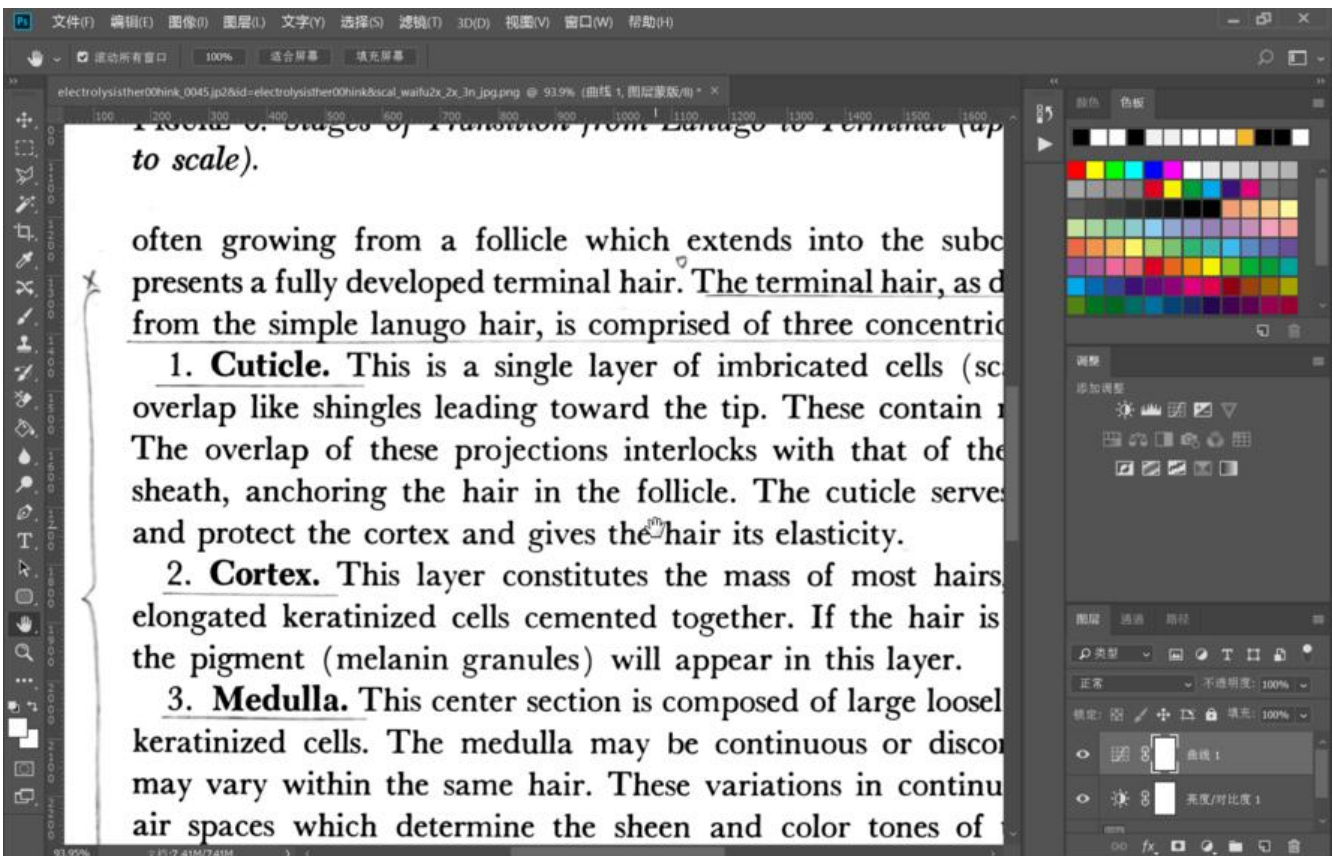
我是没有独显的笔记本，R5-4800H 处理器，纯 CPU 处理，速度大约 7 秒一张图片

看下输出前后的对比：





然后到 Photoshop 中，将图片模式转为灰度，使用曲线，调高对比度，将背景杂色进一步去掉：



合并所有图层

将 DPI 在分辨率不变下调成 144 或者更高

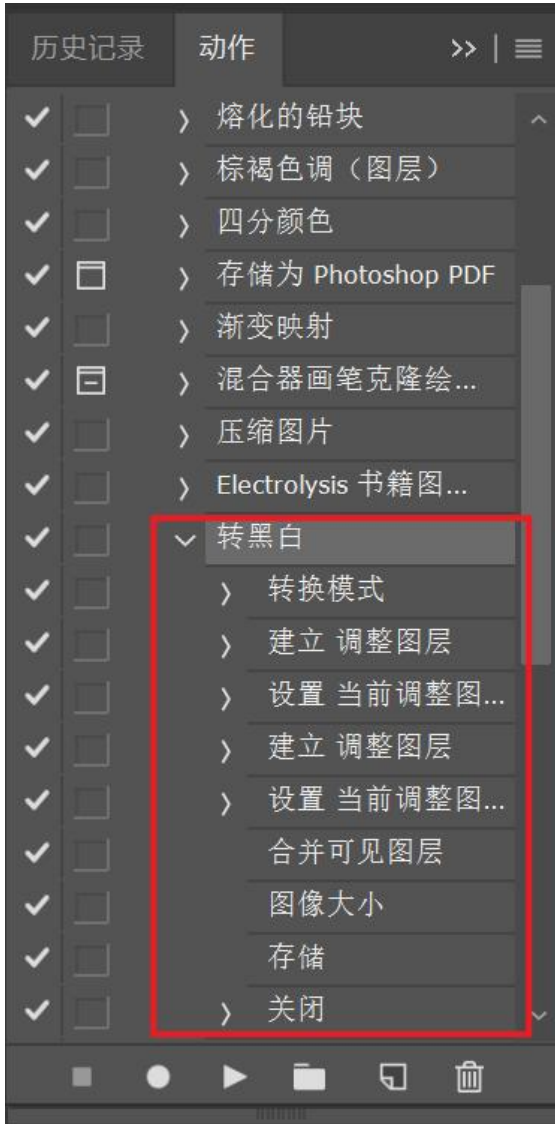
为了让 OCR 生成的字体曲线更平滑好看，我们在 Waifu2x 中，将图片进行了 2x 放大。

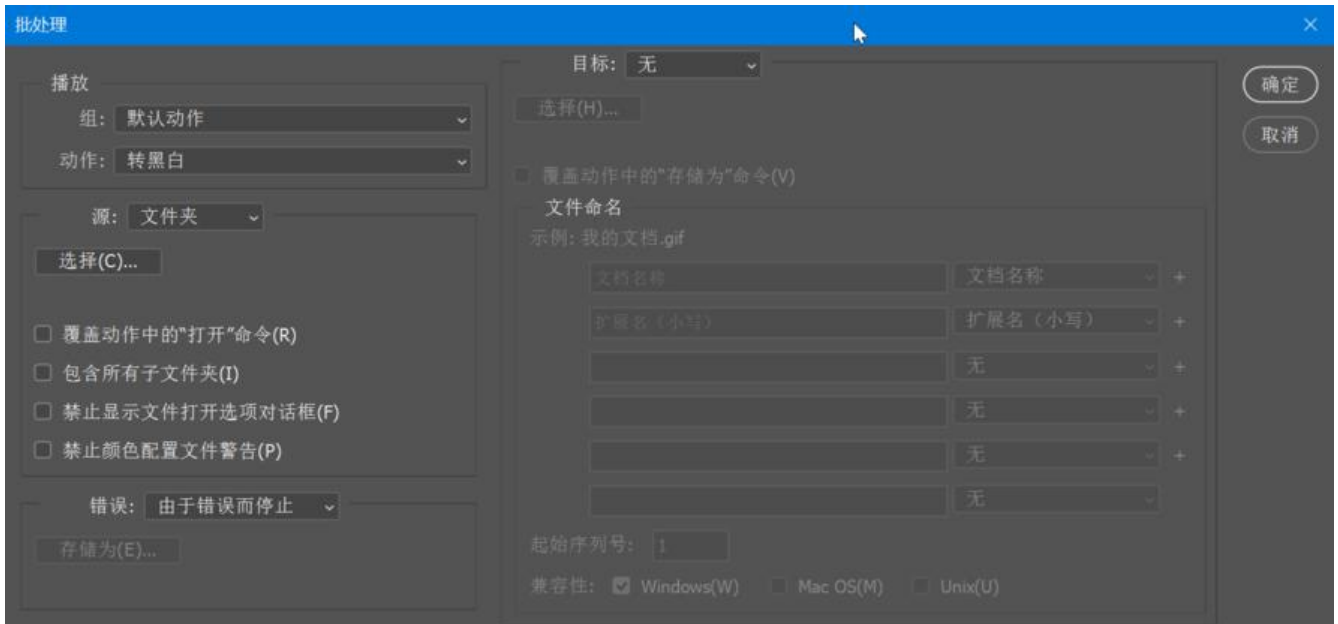
但是为了让文字之外的图片体积小些，就需要在 OCR 之后，降低图片的分辨率。

之后在 Acrobat OCR 步骤中可以调整 OCR 后图片的 DPI，例如 72。

所以提前将图片设为较高的 DPI，就可以在 OCR 步骤完成后将图片分辨率降低。减小体积占用以 100 的品质保存到原来的 JPG 文件。

将上述 PS 中的操作，在动作面板中录制为动作，对其它图片运行批处理：

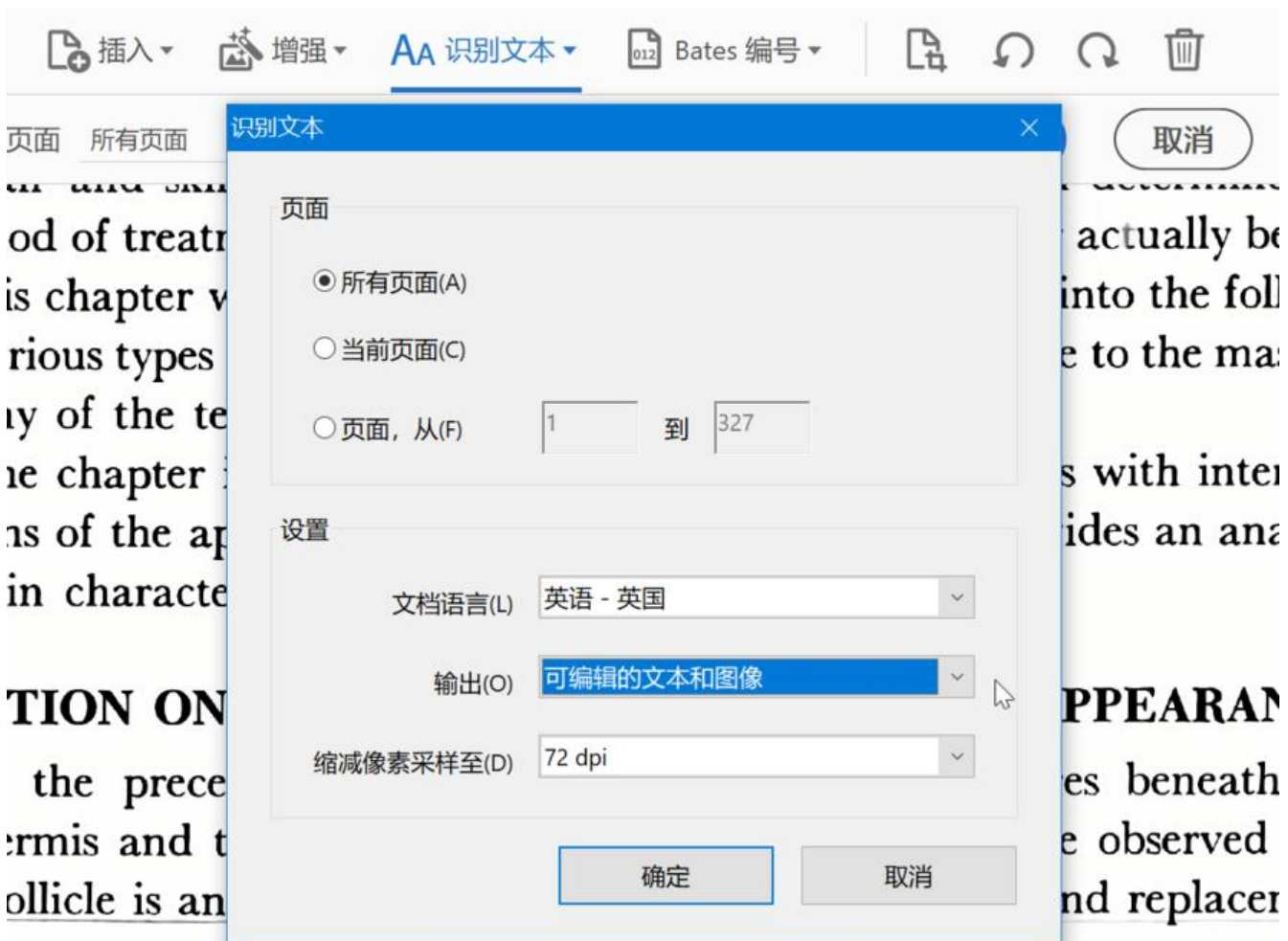




## 合并为新 PDF

将批处理后的图片全选，在 Acrobat 中合并，使用最高品质，将合并好的 PDF 保存，此时它的大小能有几个 GB。



然后使用 OCR 工具，设置如下：





然后静待识别完成，所有的文字都变成了矢量

保存为新 PDF，体积暴减：

 Electrollysis, thermolys...	23,327 KB
 Electrollysis, thermolys...	77,940 KB