## Proceedings of OCR

My script currently appears as follows:

```
#!/bin/bash
foldername="$1"
cp -av "$foldername" "$foldername"-raw
echo "Folder copied :"
ls -l "$foldername"
echo "entering..."
cd "$foldername"
echo "OCR in progress..."
for fname in *.pdf; do
 echo "#____"
 echo "Currently processing:"
 echo "$fname"
 echo "..."
 ocrmypdf -l deu -c -d -O 2 --jpeg-quality 25 --oversample 300 --threshold
"$fname" "$fname"
done
cd ..
echo "DONE!!"
exit
```

This is basically workable, as ever minimalistic in style, thus you cannot pass any nonsense as argument and expect useful results from that.

The output looks like this:

```
andrew@virtsrv:/mnt/usb/ocr-proj$ ./ocrf.sh gerstacker-gold
'gerstacker-gold' -> 'gerstacker-gold-raw/gerstacker-gold'
'gerstacker-gold/05_XVII-XX_.pdf' ->
'gerstacker-gold-raw/gerstacker-gold/05_XVII-XX_.
pdf'
'gerstacker-gold/03_IX-XII_.pdf' ->
'gerstacker-gold-raw/gerstacker-gold/03_IX-XII_.pd
f'
'gerstacker-gold/02_V-VIII_.pdf' ->
'gerstacker-gold-raw/gerstacker-gold/02_V-VIII_.pd
f'
'gerstacker-gold/06_XXI-XXIV_.pdf' ->
'gerstacker-gold-raw/gerstacker-gold/06_XXI-XXIV
_.pdf'
'gerstacker-gold/07_XXV-XXVIII_.pdf' -> 'gerstacker-gold-raw/gerstacker-
gold/07_XXV-XX
VIII_.pdf'
'gerstacker-gold/04_XIII-XVI_.pdf' ->
'gerstacker-gold-raw/gerstacker-gold/04_XIII-XVI
_.pdf'
'gerstacker-gold/01_I-IV_.pdf' -> 'gerstacker-gold-raw/gerstacker-gold/01_I-
IV_.pdf'
'gerstacker-gold/.directory' -> 'gerstacker-gold-raw/gerstacker-
gold/.directory'
Folder copied :
total 63904
-rw-r--r-- 1 andrew andrew  9875693 Apr 20 10:27 01_I-IV_.pdf
-rw-r--r-- 1 andrew andrew 11859696 Apr 20 10:27 02_V-VIII_.pdf
-rw-r--r-- 1 andrew andrew 10258277 Apr 20 10:27 03_IX-XII_.pdf
-rw-r--r-- 1 andrew andrew  8321126 Apr 20 10:27 04_XIII-XVI_.pdf
-rw-r--r-- 1 andrew andrew  7715481 Apr 20 10:27 05_XVII-XX_.pdf
-rw-r--r-- 1 andrew andrew  8416233 Apr 20 10:27 06_XXI-XXIV_.pdf
-rw-r--r-- 1 andrew andrew  8975716 Apr 20 10:27 07_XXV-XXVIII_.pdf
entering...
OCR in progress...
#____
Currently processing:
01_I-IV_.pdf
...
For best results, install the optional program 'jbig2' to use the argument
--optimize
{2,3}.
WARNING -   18: [tesseract] lots of diacritics - possibly poor OCR
  INFO - Optimize ratio: 2.12 savings: 52.8%
  INFO - Output file is a PDF/A-2B (as expected)
#____
Currently processing:
02_V-VIII_.pdf
...
For best results, install the optional program 'jbig2' to use the argument
--optimize
{2,3}.
  INFO - Optimize ratio: 2.05 savings: 51.2%
  INFO - Output file is a PDF/A-2B (as expected)
#____
Currently processing:
```

```
03_IX-XII_.pdf
...
For best results, install the optional program 'jbig2' to use the argument
--optimize
{2,3}.
  INFO -    2: [tesseract] Image too small to scale!! (2x36 vs min width of
3)
  INFO -    2: [tesseract] Line cannot be recognized!!
  INFO -    2: [tesseract] Image too small to scale!! (3x36 vs min width of
3)
  INFO -    2: [tesseract] Line cannot be recognized!!
  INFO - Optimize ratio: 2.04 savings: 51.1%
  INFO - Output file is a PDF/A-2B (as expected)
#____
Currently processing:
04_XIII-XVI_.pdf
...
For best results, install the optional program 'jbig2' to use the argument
--optimize
{2,3}.
  INFO - Optimize ratio: 2.04 savings: 51.1%
  INFO - Output file is a PDF/A-2B (as expected)
#____
Currently processing:
05_XVII-XX_.pdf
...
For best results, install the optional program 'jbig2' to use the argument
--optimize
{2,3}.
  INFO - Optimize ratio: 2.06 savings: 51.4%
  INFO - Output file is a PDF/A-2B (as expected)
#____
Currently processing:
06_XXI-XXIV_.pdf
...
For best results, install the optional program 'jbig2' to use the argument
--optimize
{2,3}.
  INFO - Optimize ratio: 2.06 savings: 51.5%
  INFO - Output file is a PDF/A-2B (as expected)
#____
Currently processing:
07_XXV-XXVIII_.pdf
...
For best results, install the optional program 'jbig2' to use the argument
--optimize
{2,3}.
WARNING -   20: [tesseract] lots of diacritics - possibly poor OCR
  INFO -   20: [tesseract] Image too small to scale!! (3x36 vs min width of
3)
  INFO -   20: [tesseract] Line cannot be recognized!!
WARNING -   21: [tesseract] lots of diacritics - possibly poor OCR
  INFO -   21: [tesseract] Image too small to scale!! (2x36 vs min width of
3)
  INFO -   21: [tesseract] Line cannot be recognized!!
  INFO - Optimize ratio: 2.05 savings: 51.3%
  INFO - Output file is a PDF/A-2B (as expected)
DONE!!
```