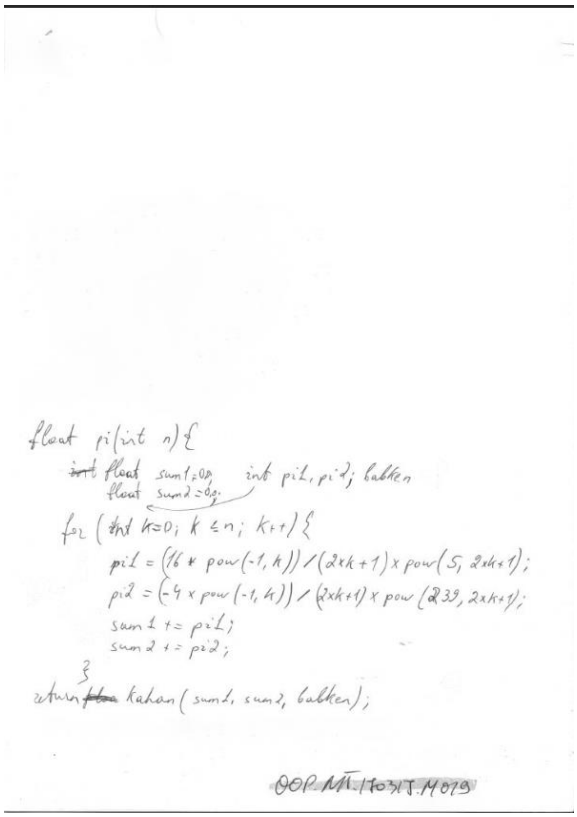
Removing the Text from the image:

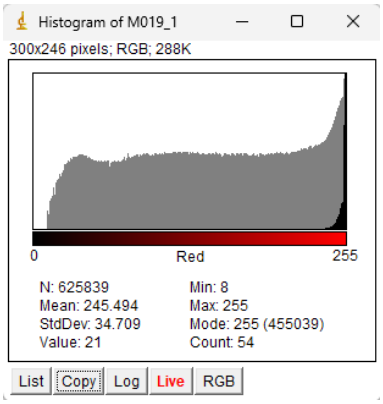
I have tried using photoshop, to remove the images. I managed to do in the following way.

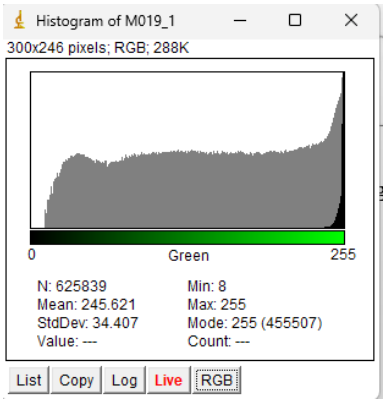
1. I selected the area of the printed text.
2. Pressed on edit at the top toolbar, then clicked on fill.
3. There's an option called contact-aware, and it is on by default.
4. After that the printed text was gone.
5. Next, I changed the contrast to black and white, and reduced the level of "red" from the RGB settings to get rid of the red colors. the result is displayed on the right.

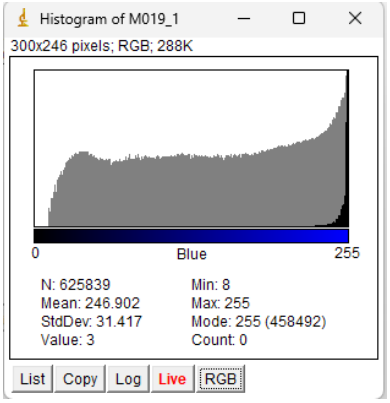


For the automation case, I have researched some tools and libraries for Python, that can be used to achieve desired outcome. Some of these tools were, *Pytesseract*, *Pillow*, *OpenCV2 and OCR (Optical Character Recognition.)*

RGB characteristics of the image with ImageJ, and I got the following results.







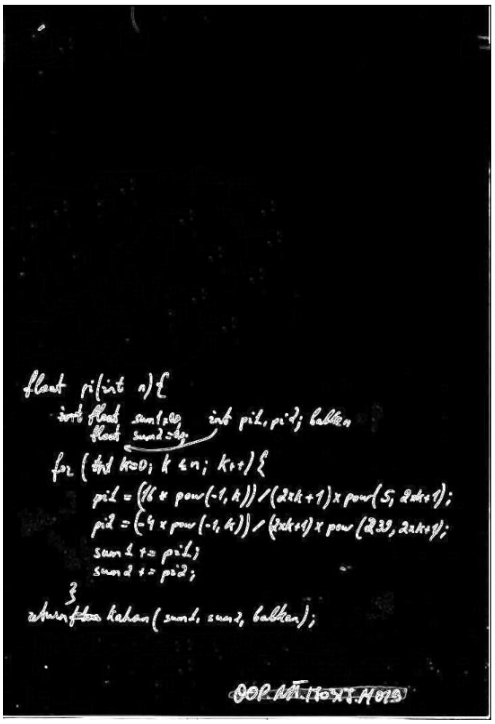
In addition, I have analyzed the

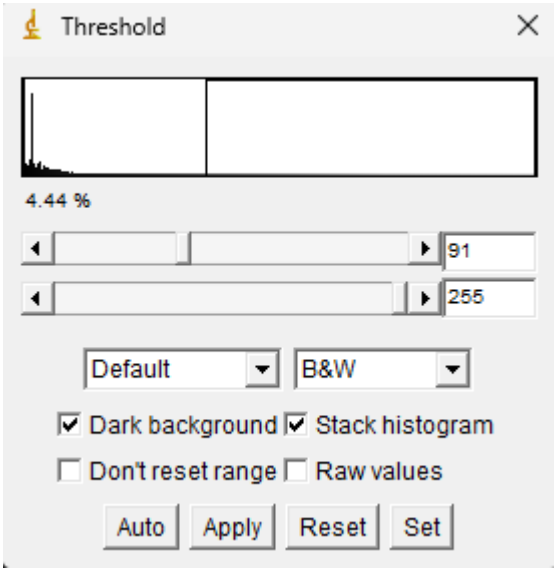
Page Features:

Since most of our exams are printed on A4 papers, therefore they're based on the standard size of the A4 (210 × 297 millimeters or 8.27 × 11.69 inches).

Stage 3: Straight Lines in Handwriting Samples:

I applied edge cutting, and after edge cutting, and setting the image in 8-bits I set the threshold to Black and White, and I set the level to the following. Thus, It gave me the picture on the right, where the handwritten text in minimized, and only the edge are remained in white.





In addition, I have discovered several tools on Github. One of them is called WordDetector. This tool has the ability to read the handwritten words, and print out the words in digital format. It is based on the research carried out by Scale Space Technique for Word Segmentation in Handwritten documents conducted by R. Manmatha and Nitin Srimal.

There is also another tool that is called:

HandwritingRecognitionSystem: Handwriting Recognition System based on a deep Convolutional Recurrent Neural Network architecture

Stage 4 – binary regions in handwriting samples

After using the tools above, we can use the MorpholibJ plugin for ImageJ to construct the binary regions for each of the cropped words.