

Contributing

Please follow these rules and advice.

Creating an Issue or Using the Forum

If you think you found a bug in Tesseract, please create an issue.

Use the [user forum](#) instead of creating an issue if ...

- You have problems using Tesseract and need some help.
- You have problems installing the software.
- You are not satisfied with the accuracy of the OCR, and want to ask how you can improve it. Note: You should first read the [ImproveQuality](#) documentation.
- You are trying to train Tesseract and you have a problem and/or want to ask a question about the training process. Note: You should first read the **official** guides [\[1\]](#) or [\[2\]](#) found in the project documentation.
- You have a general question.

An issue should only be reported if the platform you are using is one of these:

- Linux (but not a version that is more than 4 years old)
- Windows (Windows 7 or newer version)
- macOS (last 3 releases)

For older versions or other operating systems, use the Tesseract forum.

When creating an issue, please report your operating system, including its specific version: "Ubuntu 16.04", "Windows 10", "Mac OS X 10.11" etc.

Search through open and closed issues to see if similar issue has been reported already (and sometimes also has been solved).

Similarly, before you post your question in the forum, search through past threads to see if similar question has been asked already.

Read the [documentation](#) before you report your issue or ask a question in the forum.

Only report an issue in the latest official release. Optionally, try to check if the issue is not already solved in the latest snapshot in the git repository.

Make sure you are able to replicate the problem with Tesseract command line program. For external programs that use Tesseract (including wrappers and your own program, if you are developer), report the issue to the developers of that software if it's possible. You can also try to find help in the Tesseract forum.

Each version of Tesseract has its own language data you need to obtain. You **must** obtain and install trained data for English (eng) and osd. Verify that Tesseract knows about these two files (and other trained data you installed) with this command: `tesseract --list-langs`.

Post example files to demonstrate the problem. BUT don't post files with private info (about yourself or others).

When attaching a file to the issue report / forum ...

- Do not post a file larger than 20 MB.
- GitHub supports only few file name extensions like `.png` or `.txt`. If GitHub rejects your files, you can compress them using a program that can produce a zip archive and then load this zip file to GitHub.

Do not attach programs or libraries to your issues/posts.

For large files or for programs, add a link to a location where they can be downloaded (your site, Git repo, Google Drive, Dropbox etc.)

Attaching a multi-page TIFF image is useful only if you have problem with multi-page functionality, otherwise attach only one or a few single page images.

Copy the error message from the console instead of sending a screenshot of it.

Use the toolbar above the comment edit area to format your comment.

Add three backticks before and after a code sample or output of a command to format it (The `Insert code` button can help you doing it).

If your comment includes a code sample or output of a command that exceeds ~25 lines, post it as attached text file (`filename.txt`).

Use `Preview` before you send your issue. Read it again before sending.

Note that most of the people that respond to issues and answer questions are either other 'regular' users or **volunteers** developers. Please be nice to them :-)

The [tesseract developers](#) forum should be used to discuss Tesseract development: bug fixes, enhancements, add-ons for Tesseract.

Sometimes you will not get a respond to your issue or question. We apologize in advance! Please don't take it personally. There can be many reasons for this, including: time limits, no one knows the answer (at least not the ones that are available at that time) or just that your question has been asked (and has been answered) many times before...

For Developers: Creating a Pull Request

You should always make sure your changes build and run successfully.

For that, your clone needs to have all submodules (`googletest` , `test`) included. To do so, either specify `--recurse-submodules` during the initial clone, or run `git submodule update --init --recursive NAME` for each `NAME` later. If `configure` already created those directories (blocking the clone), remove them first (or `make distclean`), then clone and reconfigure.

Have a look at [the README](#) and [testing README](#) and the [documentation](#) on installation.

In short, after running `configure` from the build directory of your choice, to build the library and CLI, run `make` . To test it, run `make check` . To build the training tools, run `make training` .

As soon as your changes are building and tests are succeeding, you can publish them. If you have not already, please [fork](#) tesseract (somewhere) on GitHub, and push your changes to that fork (in a new branch). Then [submit as PR](#).

Please also keep track of reports from CI (automated build status) and Coverity/LGTM (quality scan). When the indicators show deterioration after your changes, further action may be required to improve them.