

🏠 Help the Python Software Foundation raise \$60,000 USD by December 31st!

Building the PSF Q4 Fundraiser [🔗](#)



Search projects



[Help](#)

[Sponsor](#)

[Log in](#)

[Register](#)

pdf2image 1.14.0



Latest version

`pip install pdf2image`

Released: Aug 23, 2020

A wrapper around the pdftoppm and pdftocairo command line tools to convert PDF to a PIL Image list.

Navigation

[Project description](#)

[Release history](#)

[Download files](#)

Project links

[Homepage](#)

Statistics

Project description

pdf2image

build	passing	pypi package	1.14.0	codecov	97%
downloads/month	143k	help	mattermost	docs	passing

A python (3.5+) module that wraps pdftoppm and pdftocairo to convert PDF to a PIL Image object

How to install

```
pip install pdf2image
```

Windows

🔗 **Help the Python Software Foundation raise \$60,000 USD by December 31st!**

Building the PSF Q4 Fundraiser 🔗

! Open

issues/PRs: 32

View statistics for this project via

[Libraries.io](#) 🔗, or by using [our public dataset on Google BigQuery](#) 🔗

Meta

License: MIT License (MIT)

Author: [Edouard Belval](#) ✉

📁 pdf, image, png, jpeg, jpg, convert

Maintainers



Belval

Classifiers

Development Status

- 5 - Production/Stable

Intended Audience

- Developers

License

- OSI Approved :: MIT License

```
poppler_path = C:\path\to\poppler\bin as an
argument in convert_from_path.
```

Mac

Mac users will have to install [poppler for Mac](#).

Linux

Most distros ship with `pdftoppm` and `pdftocairo`. If they are not installed, refer to your package manager to install `poppler-utils`

Platform-independant (Using `conda`)

1. Install poppler: `conda install -c conda-forge poppler`
2. Install pdf2image: `pip install pdf2image`

How does it work?

```
from pdf2image import convert_from_path,
convert_from_bytes
```

```
from pdf2image.exceptions import (
    PDFInfoNotInstalledError,
    PDFPageCountError,
    PDFSyntaxError
)
```

Then simply do:

```
images = convert_from_path('/home/belval/example.pdf')
```

OR

🏠 Help the Python Software Foundation raise \$60,000 USD by December 31st!

[Building the PSF Q4 Fundraiser](#)

- [Python :: 3.6](#)
- [Python :: 3.7](#)
- [Python :: 3.8](#)

OR better yet

```
import tempfile

with tempfile.TemporaryDirectory() as path:
    images_from_path = convert_from_path('/home/belval/ex
    # Do something here
```

`images` will be a list of PIL Image representing each page of the PDF document.

Here are the definitions:

```
convert_from_path(pdf_path, dpi=200, output_folder=None,
first_page=None, last_page=None, fmt='ppm', jpegopt=None,
thread_count=1, userpw=None, use_cropbox=False,
strict=False, transparent=False, single_file=False,
output_file=str(uuid.uuid4()), poppler_path=None,
grayscale=False, size=None, paths_only=False,
use_pdftocairo=False, timeout=600)
```

```
convert_from_bytes(pdf_file, dpi=200, output_folder=None,
first_page=None, last_page=None, fmt='ppm', jpegopt=None,
thread_count=1, userpw=None, use_cropbox=False,
strict=False, transparent=False, single_file=False,
output_file=str(uuid.uuid4()), poppler_path=None,
grayscale=False, size=None, paths_only=False,
use_pdftocairo=False, timeout=600)
```

Need help?

Use the [mattermost chat](#) to ask questions on the helpdesk and get direct support.

What's new?

🏠 Help the Python Software Foundation raise \$60,000 USD by December 31st!

Building the PSF Q4 Fundraiser [🔗](#)

`pdftocairo`. Should improve performance.

- Fixed a bug where using `pdf2image` with multiple threads (but not multiple processes) would cause an exception
- `jpegopt` parameter allows for tuning of the output JPEG when using `fmt="jpeg"` (`-jpegopt` in `pdftoppm` CLI) (Thank you @abieler)
- `pdfinfo_from_path` and `pdfinfo_from_bytes` which expose the output of the `pdfinfo` CLI
- `paths_only` parameter will return image paths instead of Image objects, to prevent OOM when converting a big PDF
- `size` parameter allows you to define the shape of the resulting images (`-scale-to` in `pdftoppm` CLI)
 - `size=400` will fit the image to a 400x400 box, preserving aspect ratio
 - `size=(400, None)` will make the image 400 pixels wide, preserving aspect ratio
 - `size=(500, 500)` will resize the image to 500x500 pixels, not preserving aspect ratio
- `grayscale` parameter allows you to convert images to grayscale (`-gray` in `pdftoppm` CLI)
- `single_file` parameter allows you to convert the first PDF page only, without adding digits at the end of the `output_file`
- Allow the user to specify poppler's installation path with `poppler_path`

Performance tips

- Using an output folder is significantly faster if you are using an SSD. Otherwise i/o usually becomes the bottleneck.
- Using multiple threads can give you some gains but avoid more than 4 as this will cause i/o bottleneck (even on my NVMe SSD!).
- If i/o is your bottleneck, using the JPEG format can lead to significant gains.
- PNG format is pretty slow, this is because of the compression.

🏠 **Help the Python Software Foundation raise \$60,000 USD by December 31st!**

Building the PSF Q4 Fundraiser [🔗](#)

Limitations / known issues

- A relatively big PDF will use up all your memory and cause the process to be killed (unless you use an output folder)



Help

Installing packages [🔗](#)

Uploading packages [🔗](#)

User guide [🔗](#)

FAQs

About PyPI

PyPI on Twitter [🔗](#)

Infrastructure dashboard [🔗](#)

Package index name retention [🔗](#)

Our sponsors

Contributing to PyPI

Bugs and feedback

Contribute on GitHub [🔗](#)

Translate PyPI [🔗](#)

Development credits [🔗](#)

Using PyPI

Code of conduct [🔗](#)

Report security issue

Privacy policy [🔗](#)

Terms of use

Status: All Systems Operational [🔗](#)

Developed and maintained by the Python community, for the Python community.

Donate today!

© 2020 Python Software Foundation [🔗](#)

Site map

🏠 Help the Python Software Foundation raise \$60,000 USD by December 31st!

[Building the PSF Q4 Fundraiser](#)

► English español français 日本語 português (Brasil) українська Ελληνικά Deutsch 中文 (简体) русский
עברית

Pingdom
Monitoring

Google
Object Storage and
Download Analytics

Sentry
Error logging

AWS
Cloud computing

DataDog
Monitoring

Fastly
CDN

DigiCert
EV certificate

StatusPage
Status page