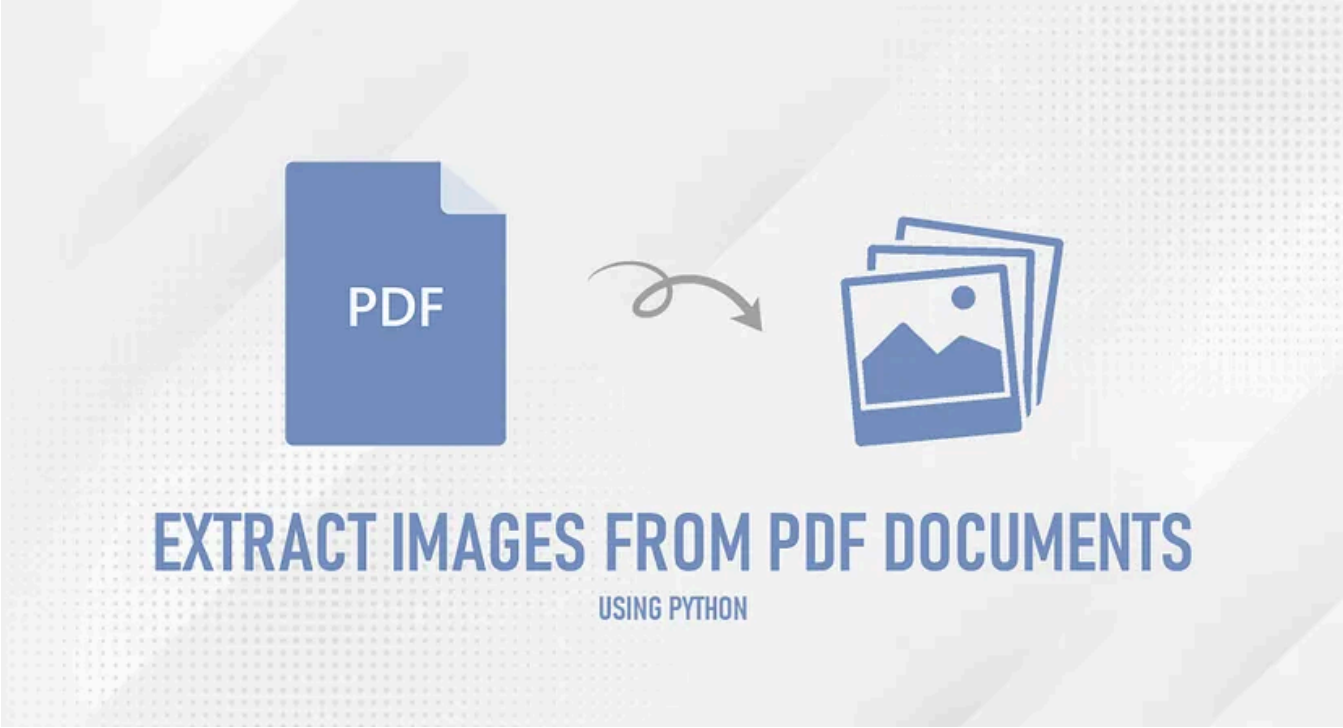
Extract Images from PDF Documents in Python

Alexander Stock · Follow
3 min read · Oct 12, 2023



Extracting images from a PDF file can be a useful and practical task in various situations. Whether you need to repurpose images for a presentation, create a digital photo album, or simply save images for future reference, the ability to extract images from a PDF can save you time and effort. In this article, you will learn how to **extract images from a PDF document in Python** using Spire.PDF for Python.

- Extract Images from a Specific Page in Python
- Extract All Images from a PDF Document in Python

Install Dependency

This solution requires Spire.PDF for Python to be installed as the dependency, which is a Python library for reading, creating and manipulating PDF documents in a Python program. You can install it by running the following pip command.

```
pip install Spire.PDF
```

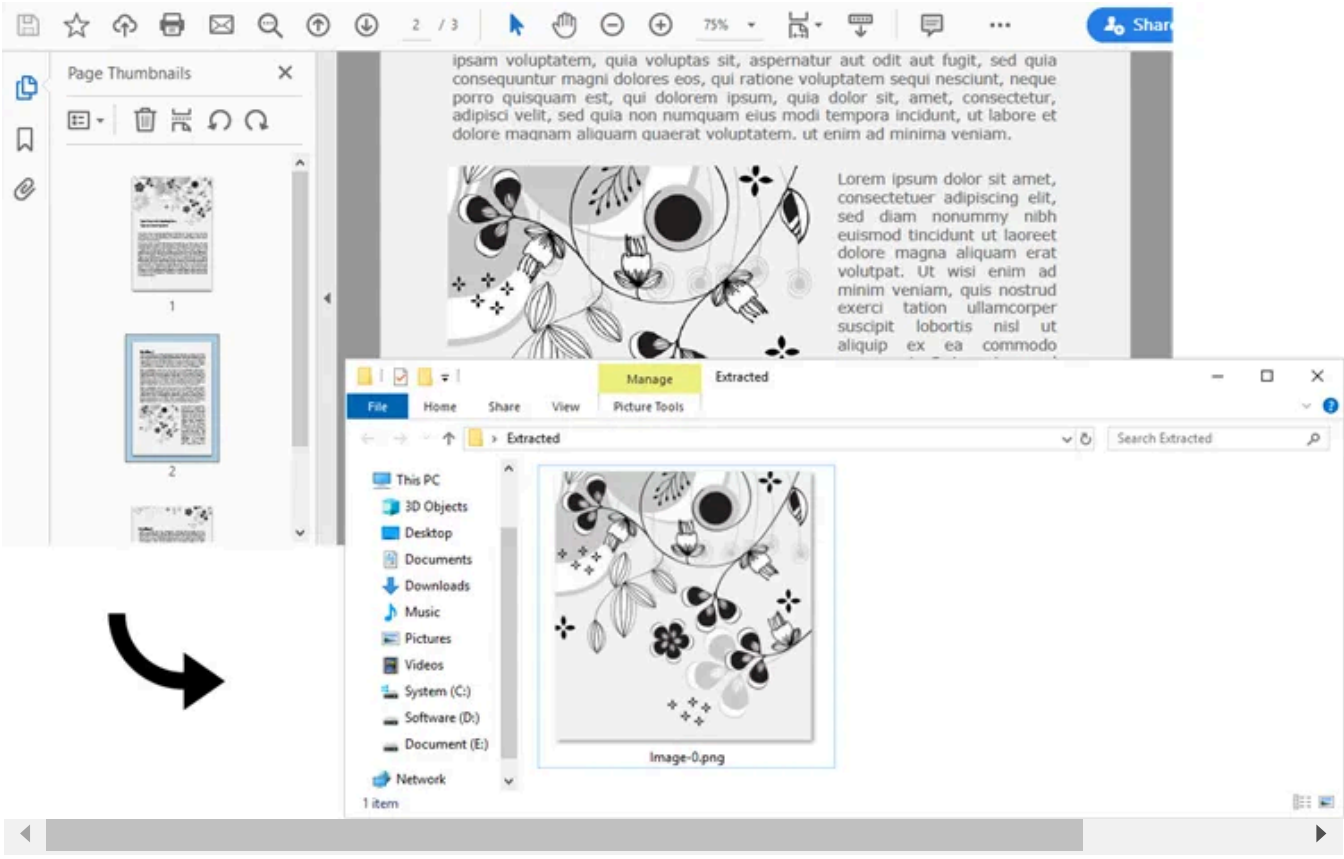
Extract Images from a Specific Page in Python

Spire.PDF for Python offers the **PdfPageBase.ExtractImages()** method to extract images from a specified page. The following are the detailed steps.

- Create a **PdfDocument** object.
- Load a PDF document using **PdfDocument.LoadFromFile()** method.
- Get a particular page through **PdfDocument.Pages[index]** property.
- Extract images from the page using **PdfPageBase.ExtractImages()** method and return a list of images.
- Write each image in the list as a PNG file.

```
from spire.pdf.common import *
from spire.pdf import *

# Create a PdfDocument object
doc = PdfDocument()

# Load a PDF document
doc.LoadFromFile('C:/Users/Administrator/Desktop/input.pdf')

# Get a specific page
page = doc.Pages[1]

# Extract images from the page
images = []
for image in page.ExtractImages():
    images.append(image)

# Save images to specified location with specified format extension
index = 0
for image in images:
    imageFileName = 'C:/Users/Administrator/Desktop/Extracted/Image-{0:d}.png'.
    index += 1
```

```
    image.Save(imageFileName, ImageFormat.get_Png())
doc.Close()
```



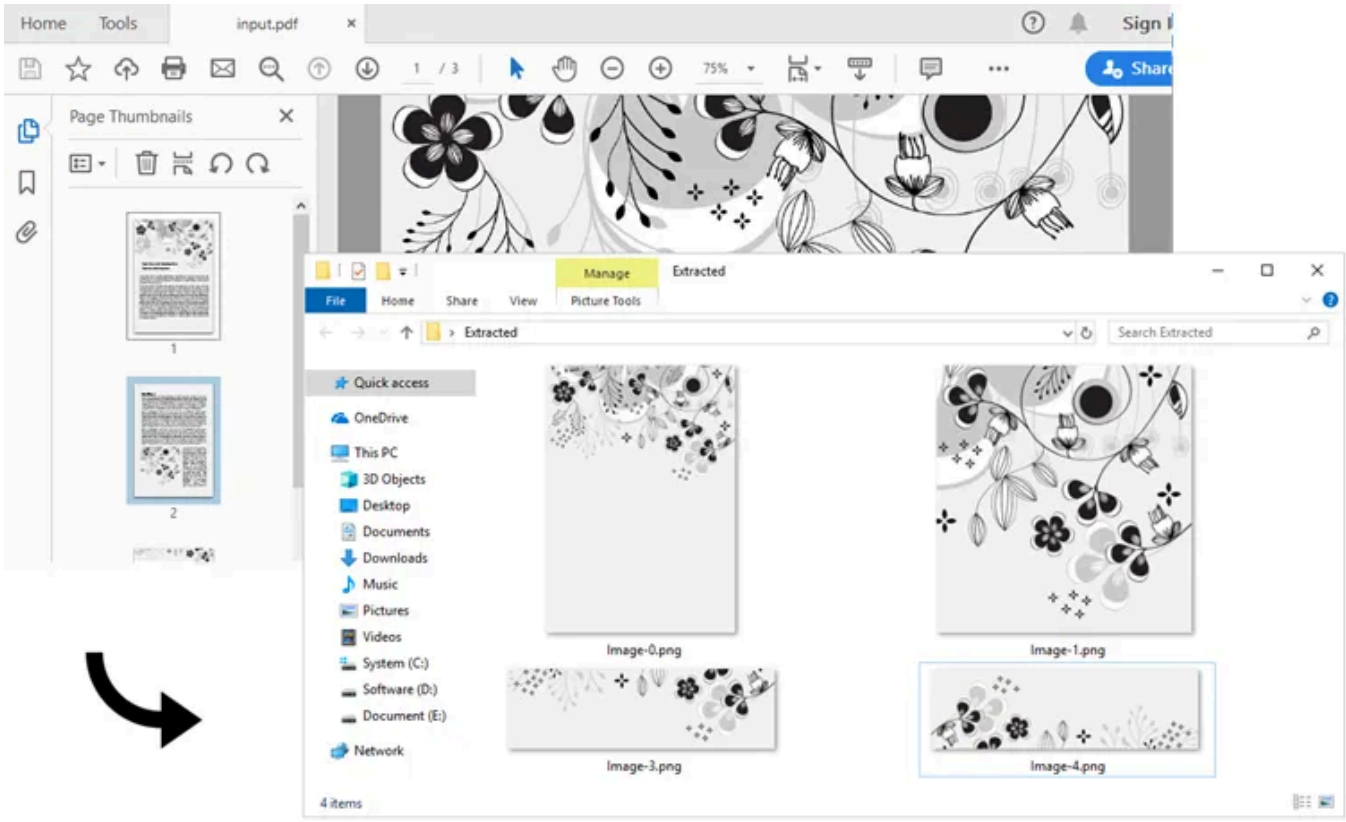
Extract All Images from a PDF Document in Python

To extract all images from an entire PDF document, loop through the pages in the document and then retrieve the images from each page separately. Here are the detailed steps.

- Create a **PdfDocument** object.
- Load a PDF document using **PdfDocument.LoadFromFile()** method.
- Iterate through the pages in the document, and get the images from each page using **PdfPageBase.ExtractImages()** method.
- Write all extracted images as individual PNG files.

```
from spire.pdf.common import *
from spire.pdf import *

# Create a PdfDocument object
doc = PdfDocument()
```

```
# Load a PDF document
doc.LoadFromFile('C:/Users/Administrator/Desktop/input.pdf')

images = []

# Loop through the pages in the document
for i in range(doc.Pages.Count):
    page = doc.Pages.get_Item(i)

    # Extract images from a specific page
    for image in page.ExtractImages():
        images.append(image)

# Save images to specified location with specified format extension
index = 0
for image in images:
    imageFileName = 'C:/Users/Administrator/Desktop/Extracted/Image-{0:d}.png'.
    index += 1
    image.Save(imageFileName, ImageFormat.get_Png())
doc.Close()
```



Conclusion

This blog post provides valuable insights into extracting images from PDF documents using Python. It covers two main techniques: extracting images from a specific page and extracting all images from a PDF document. These techniques