PyPI: an internet-enabled learning tool to boost learner motivation for pest science

Yoshiyasu Takefuji (✉️ takefuji@keio.jp)
Musashino University

Research Article

Keywords: counting disaggregated objects, PyPI, software dissemination, learner's incentive

Posted Date: August 21st, 2023

DOI: https://doi.org/10.21203/rs.3.rs-3251122/v1

License: ☑️ Ⓢ This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License
Abstract

Instructors are always interested in methods to activate learner incentives and motivation to increase learning effectiveness. This paper introduces the Python Package Index (PyPI) as a powerful tool to maximize learner incentives on software and presents an example of its application in pest science. The more useful the PyPI application is, the more it will be downloaded worldwide, providing an external review for the learner and strengthening their incentive. However, many existing tutorials on PyPI, including the official site, are not updated on the twine library for uploading files to the PyPI site. This paper presents an updated tutorial on using PyPI for counting disaggregated objects such as bugs and pests, and for software reproducibility validation via Code Ocean. Additionally, generative AI is introduced as an indispensable assistant for tasks such as understanding technical terms and providing solutions for encountered problems.

Practitioner Notes

What is already known about this topic

• Instructors don't know how to maximize motivation and incentives of learners.
• Python is widely used in science and technology, including pest science.
• Instructors have been looking for the best practice to boost the learner's motivation.

What this paper adds

• Internet-enable learning with PyPI can maximize motivation of learners on software.
• The third party evaluation plays a key role in maximizing the learner’s motivation.
• The software reproducibility validation is conducted via Code Ocean.

• This may be the first tutorial on software dissemination in pest science and technology.

Implications for practice and/or policy

• Instructors must learn how to debut a PyPI application to disseminate it to the world.
• The more the downloads, the stronger the motivation for learners.
• Disseminating PyPI applications can show skills of learners for possible recruitment.

• Software reproducibility validation via Code Ocean is introduced for pest scientists.

Highlights

• This paper presents how to debut PyPI applications for pest scientists.
• The bugcount PyPI can count the number of disaggregated bugs and pests.
• PyPI is an internet-enabled learning powerful tool to stimulate learner's incentives.
• Software reproducibility validation via Code Ocean for PyPI app is introduced.

INTRODUCTION
In pest science, understanding the pest is crucial. This includes knowledge of its biology, behavior, ecology, and interactions with other organisms. With this understanding, we can develop effective and sustainable pest management strategies. This paper presents a new application for pest scientists, encouraging novice and non-programmers to develop and share their prototype tools with the world. This provides motivation and incentives for the development and dissemination of innovative pest management solutions.

Instructors are always interested in methods on how to improve learner incentives to increase learning effectiveness regardless of research areas. The purpose of this paper is not only to introduce the Python Package Index (PyPI) to maximize learner incentives in higher education and industry education on software, but also to present an example of PyPI applications. The more useful it is, the more its PyPI application will be downloaded worldwide. The more downloads of the PyPI-application, the stronger the incentive for learners on software. PyPI is a powerful internet-enabled tool that helps learners increase their motivation and incentives.

This paper demonstrates an example of how to debut a Python Package Index (PyPI) application for learners in pest science. The more useful it is, the more its PyPI application will be downloaded worldwide. The proposed example of the PyPI application is to count the number of disaggregated objects such as cells, pests, flies, pollens, dead insect bodies or bug-corpses on a plane, which is widely used in academia and industry. This example is intentionally simple, but powerful for the purposes of this paper for pest scientists.

Machine learning has been successfully applied to count the number of objects [1, 2, 3, 4]. However, one of the challenges of using machine learning for this task is the need to collect and annotate large datasets, which can be a very time-consuming process. This paper will show the tunable Canny edge detection method with the Gaussian Blur function for noise elimination without machine learning. In other words, the time-consuming annotation is not needed with the proposed method.

According to PePy: https://pepy.tech/, the proposed software, bugcount has been downloaded by 12344 times worldwide. The large number of downloads of bugcount indicates the popularity. Remember that the more downloads, the stronger the motivation or incentives for learners. The bugcount PyPI is one of the most popular applications of its kind in all areas including pest science.

PyPI has many tutorials on the Internet, but even the official PyPI site has not been updated about uploading files to the PyPI site using the twine library. Following them will fail to upload the PyPI package due to twine's new library changes. To our knowledge, there is no up-to-date tutorial on how to debut a PyPI application.

There are no articles on updated PyPI tutorial on the twine library in the existing tutorials. Thus, the proposed tutorial is significant for advancing science and technology on software.
Learner incentives play an important role in general skill development. Software teachers and instructors are always interested in techniques to activate learner motivation and incentives. The more downloads of PyPI application, the more incentives given to learners. When an outsider in the world who does not know a learner but downloads his/her PyPI application, it is seen as an external review for the learner, so if many downloads occur, it can be judged as an external review very good, thus activating the learner's incentives and motivation.

A numerical blob method was used for a range of diffusive partial differential equations of Wasserstein gradient flow type [5]. Blob stands for binary large object which is a collection of binary data stored as a single entity. The term “large” indicates that only objects of a certain size are of interest and that “small” binary objects are usually noise. Thus, blobs are images in this paper.

The proposed tunable blob method is combined with a Gaussian blur function to eliminate noise for use in real-world applications for counting the number of disaggregated objects. The advantage of the proposed method is its tunable ability to allow users to achieve perfect counting. The disadvantage of the proposed method is that the user needs to do the tuning. This tuning is similar to how many instances of a dataset a user has to prepare in machine learning.

For a large dataset, GPUs are needed in machine learning. The cost of GPU technology has been dramatically reduced, but GPU computers are still expensive compared to traditional CPU computers. However, applications that can be run on CPU computers without GPU hardware are still highly demanded. The proposed real-time algorithm in this manuscript is based on CPU machines. The second advantage of the proposed method is that it does not require a GPU.

The real-time CPU algorithm means that it runs not only on CPU machines but also on GPU machines.

The conventional counting algorithms using machine learning are based on object detection and object segmentation. Therefore, a large dataset of images for objects detection and objects segmentation is needed in machine learning. Also, collecting and annotating images is a tedious and time-consuming task, while huge datasets of images must be prepared and fed to expensive GPU machines.

In the worst-case, in counting the number of dead insects, a variety of oriented dead images must be used. Besides, deformed corpses (dead bodies) with varying degrees of dryness have too many shapes. In other words, a variety of shapes of objects must be used for correct object detection and segmentation using the conventional machine learning algorithms. In this paper, we do not consider the classification of objects. The conventional machine learning requires the collection of object shapes with various orientations to achieve the goal of accurate counting.

The proposed method is a simplest method without machine learning where the number of disaggregated objects can be counted by using the concept of blob. Blob refers to a group of connected pixels in a binary image [6, 7].
Selecting corpses or a piece of corpses can be determined by the size of object images. The proposed method has a tuning capability with a Canny coefficient. The tunable Canny coefficient allows users to determine the size of object images.

Xiaomin Guo et al. proposed a cell-counting blob method [8]. However, their program specializes in counting cells and does not use open-source software, so this method cannot be used for other purposes and applications.

Yoshinaga et al. presented a blob descriptor approach for people counting [9]. However, their approach uses machine learning segmentation. In other words, their method needs tedious machine learning.

Stajnko D. et al. proposed a blob method for counting apple fruits [10]. However, their method is based on the colored features such as red color so that their method can only be used for counting colored objects.

The proposed method is based on blobs without color features so that it can be applied to bug-counting, cell-counting, and other counting disaggregated objects.

As far as we know the existing blob methods for counting disaggregated objects, the proposed program, bugcount is one and only one open-source software with the tunable capability which can be installed by de facto standard of PyPI packaging.

In the proposed method, unnecessary noises can be eliminated by repetitive Gaussian Blur functions. Canny edge detection is used for segmenting target objects. The tunable Canny coefficient in the propose method allows users to determine the size of object images. As far as we know, there is no tunable open-source software for counting disaggregated images which can run on the conventional inexpensive CPU machines.

Canny edge detection and Gaussian blur function are available in OpenCV library of Python based on partial differential equations. The disclosed Python program will run on a notebook PC without a GPU.

The proposed application is one and only one open-source real-time software that runs lightly on a CPU computer. The proposed method has advantages over the existing methods for counting disaggregated objects, which is widely used in industry. However, as mentioned earlier, object classification is not possible with the proposed method.

This paper will show how to debut the packaging software application PyPI to maximize software dissemination in private and business settings. Only a few papers have shown how to debut the PyPI package [11, 12, 13]. PyPI packaging needs three files: README.md, setup.py, and bugcount.py. The PyPI packaging will be detailed in this paper.

In traditional software development, programmers must write programs from scratch. Therefore, a lot of time is required for programming. Rapid open-source prototyping can complete the target program within a few hours. However, selecting the right libraries requires specialized skills.
The contribution of this paper is to select the right library such as OpenCV which is one of the best image processing libraries. However, OpenCV is a complex library so that selecting sub-libraries such as plays a key role in the quality of results. However, in this paper, bugcount allows users to tune two parameters such as Gaussian Blur functions and Canny edge detection by themselves for achieving the satisfiable goal.

In the past, learners had to manually modify Python code and seek help from human experts for unfamiliar tasks in programming and installation. With the advent of generative AI, users can now interact with AI systems to solve problems and manage tasks by themselves.

This tutorial is designed for novice programmers and learners who are new to Python programming. Generative AI systems via web browser such as OpenAI's ChatGPT, Microsoft's Bing.com with ChatGPT-4, and Google's Bard for free of charge can help novice programmers and learners modify code given in this paper for their desired applications. In other words, by providing Python code to a generative AI system, it can automatically convert it to the desired application through interaction with the user, eliminating the need for manual modification. To use a generative AI system, you need to know how to access it, how to provide code via a web browser, and how to use terminal commands to run the modified program. Before running the Python program, you must install Python on your system.

To install Python package, download one of packages depending on your system:

https://docs.conda.io/en/latest/miniconda.html

It is recommended to download Python 3.9 or 3.8 instead of Python3.10, as generative AI systems may be more familiar with these older versions.

The following access links show three generative AI systems:

ChatGPT-3 via any browser: https://chat.openai.com/

Bing.com with ChatGPT-4 via Edge browser: https://bing.com/chat

Bard via any browser: https://bard.google.com/

If you are not familiar with Python installation, terminal commands, or code editing, you can ask a generative AI system for assistance. It can help with tasks such as running code, installing libraries, and modifying source code. In other words, with the help of generative AI, you no longer need to rely on human experts to solve problems or debug code. The author recommends bing.com with ChatGPT-4 and Bard.

This paper also introduces software reproducibility validation via Code Ocean. The necessary steps are detailed in this paper.
Query to generative AI

The query is crucial in obtaining solutions from generative AI. To achieve their desired goals, users need to possess three skills to effectively navigate generative AI systems:

1. the ability to interact with generative AI using appropriate language and phrasing,
2. the ability to verify the correctness of code generated by generative AI, and
3. The ability to design the code.

The first two skills mentioned in this paper are essential for users. If you encounter errors in the modified code, you can provide them to the generative AI system for correction. You can also ask the AI for details about specific lines of code.

If you are unfamiliar with technical terms or jargon, you can ask a generative AI system for clarification. Current generative AI systems have a limited number of tokens. If you exceed the maximum number of queries allowed via a web browser, the system may stop responding.

One of undergraduate students published a paper on COVID-19 policy outcome analysis tool, jpcovid [14]. The jpcovid tool has been downloaded 2,796 times which strongly motivated the undergraduate students in programming.

How to install and run bugcount

Before explaining the details of bugcount [15], the following installation procedure shows how to install bugcount via PyPI packaging installation on the system. A PyPI packaging application needs to run Python program. Therefore, you need to install the recommended Python3.8 on your machine. bugcount can run on Windows, MacOS, and Linux operating systems respectively.

1. Download the miniconda from the following site:

https://docs.conda.io/en/latest/miniconda.html

You should choose the right file depending on your operating system.

1. Install a Python running environment.

For Windows, double-click the downloaded file.

For MacOS, run the following command. ($) sign indicates the prompt from the system in command line terminal.

$ bash Miniconda3-py38_4.11.0-MacOSX-x86_64.sh

For WSL on Windows, or Linux operating systems,
$ bash Miniconda3-py38_4.11.0-Linux-x86_64.sh

1. Run the following command to install PyPI packages such as opencv and bugcount by pip command:

$ pip install opencv-python

$ pip install bugcount

To run bugcount, you should download a test image:

pillbug.png

Run bugcount by the following command:

$ bugcount pillbug.png

If you will have a trouble to show and pop the result on your screen. You may need to install X-server. For Windows, you should download VcXsrv, Windows X Server exe file and install it:
https://sourceforge.net/projects/vcxsrv/

For MacOS, you should install XQuartz

Methods

The proposed method is based on the concept of blob. Blob stands for binary large object and refers to a group of connected pixels in a binary image.

The proposed blob method with a tunable capability allows users to achieve perfect counting disaggregated objects without machine learning. The blob method runs on CPU machines that do not have a GPU component. This paper shows a single example of pill bugs with its tuning capability. Supplemental document has two examples.

The proposed method without machine learning is based on tunable Canny edge detection for counting the number of blobs while unnecessary noises can be eliminated by repetitive Gaussian Blur functions. By using Canny edge detection, contours or blobs can be detected and the number of blobs is calculated. In other words, the minimum size of blobs can be determined by a Canny coefficient. By tuning the coefficient, the proposed method can achieve perfect counting of disaggregated objects.

In order to understand the basic algorithm, the program is basically composed of the following six steps:

1. An original RGB color image is read by the Python program.
2. The original RGB color image is converted to grayscale monochrome.
3. The monochrome image is blurred several times by Gaussian Blur function.
4. Canny edge detection is applied to the monochrome image for creating contours (blobs).
5. Counting the number of contours (blobs) formed by detected Canny edges in the monochrome image.
6. Save the result.

Image segmentation using watershed algorithm with Canny edge detection and Gaussian Blur functions plays a key role in segmenting blobs in a given image.

The problem of Canny edge detection sometimes lies in that any overlapped blobs can be counted as one blob. Therefore, all overlapped-image blobs must be separated or isolated in the target picture on a single plane. In other words, all the objects to be counted need to be disaggregated.

Results

The proposed method is based on mathematical algorithms using Canny edge detection and Gaussian blur function. This Section empirically explains two functions.

Figure 1 shows four pictures (original, gray, blurred, and edged) generated by a Python program for counting the number of dead pill bugs. The function of OpenCV Gaussian blur function, cv2.GaussianBlur plays a key role in eliminating cumbersome noises. In the state-of-the-art fine-tuned blob method, the Gaussian Blur function (cv2.GaussianBlur) is called three times in the Python program for eliminating unnecessary noises.

In the fine-tuned program, Gaussian Blur plays a key role in mitigating nuisance noises by blurring an image by Gaussian function.

For novice users running Python programs, a new repository in github site was created for learners:

https://github.com/XXX/counting-for-learners

The new repository shows the followings:

1. how to install an executable Python environment on operating system (Windows, Mac, Linux),
2. how to run bugcount program,
3. how to obtain the final result using three testing pictures (pillbug.png, flies.png, and cells.png),
4. how to resize a picture and how to tune Canny edge detection coefficient.
5. Two other examples including flies counting and cell-counting are tested.

The following command shows and generates the result of five pictures (original, gray, blurred, edged and result).

As mentioned earlier, the proposed method has a tunable Canny coefficient to select blobs of a given image size. This advantageous tunable feature can distinguish between proposed and existing methods. After running the proposed program a few times, the user will be able to recognize the appropriate Canny coefficient that determines the size of a blob.
In this example, an image file of pillbug.png is used for demonstrating the proposed method. Image file and Canny coefficient can be given to bugcount where default Canny coefficient is 75:

$ bugcount pillbug.png

Figure 1 shows the result of four pictures (original: orig.png, gray: gray.png, blurred: blur.png, and edged: edge.png). Figure 1 shows three processes: a process converting a color image into a gray-scaled image, a process blurring a gray-scaled image, and a process of Canny image edging for forming contours (blobs).

———INSERT Fig. 1 ————

Figure 2 shows the final result picture where 53 dead corpses are counted in a pillbug.png file by running bugcount.

———INSERT Fig. 2 ————

If we set the Canny coefficient to 10, we can detect more blobs, as shown in Fig. 3.

$ bugcount pillbug.png 10

———INSERT Fig. 3 ————

When the Canny coefficient is set to 100, a small number of blobs are detected, as shown in Fig. 4.

$ bugcount pillbug.png 100

———INSERT Fig. 4———

**Software reproducibility validation via Code Ocean**

Refereed journals may ask authors to validate the reproducibility of the proposed software as requirement. Code Ocean is one of publishing the reproducibility badge for free of charge. You must create a new account on Code Ocean: https://codeocean.com/

The following steps are needed for software reproducibility validation:

1. Click New Capsule.
2. Create Blank Capsule.
3. In Environment, click the necessary buttons for installing Python Libraries. In this example, pick Python3.8.5 and pip3. And install matplotlib, opencv-python, and bugcount by add button.
4. Click Start with Sample Files.
5. Fill the form of metadata.
6. Modify run file
7. Click "Submit for publication" and check all items.
8. After receiving acceptance from Code Ocean after few days, you can put the badge on reproducibility qualification in the following form in any markdown document in GitHub site.

**DISCUSSION**

This paper addresses how to modify the proposed program for accommodating new counting applications.

In open-source software, source codes play a key role in accurately understanding the algorithm. The source codes are a collection of structured mathematical expressions with no ambiguity in any computer languages. Therefore, the minimum important raw Python code is explained in a plain English.

In bugcount.py Python program, outline = cv2.Canny(blurred, 0, 75*coeff) is important for counting blobs in a given image where ‘75’ is the default number to be calibrated for accommodating new counting applications. In Canny edge detection, the lower the number, the more blobs are detected. In other words, the minimum size of blobs can be determined by the Canny coefficient. This means that setting the larger Canny coefficient, a smaller number of blobs will be detected.

Another important parameter is the size of the original image. The input file (pillbug.png) is composed of PNG image data, 550 x 540, 8-bit/color RGB. However, this file was resized from the original file (p.jpg) with 4032x3024 by the following resize program:

$ python resize.py p.jpg

In order to count the number of objects using the proposed blob method without machine learning, Canny edge detection coefficient parameter and the file size must be calibrated for accommodating any counting applications. Remember that aggregated blob images cannot be distinguished in Canny edge detection so that aggregated objects in the original image must be disaggregated.

The proposed Python program will be useful for many counting applications in a variety of applications. Supplemental document contains two examples including flies.png and cells.png. Two examples can be only used for review, since they are all copyrighted.

**PyPI packaging**

To debut a PyPI packaging, three files such as README.md, setup.py, and bugcount.py are needed. README.md file can be easily created by GitHub site with creating a new repository with a README file option. To create a new account on GitHub, go to https://github.com/join

You need to create a new account on pypi site: https://pypi.org/account/register

Assume you have created a new repository for your project with GitHub and PyPI account.

1. Create a new directory of bugcount
$ mkdir bugcount

$ cd bugcount

2. Download README.md file from your GitHub site. This is an example.

$ wget https://github.com/XXX/counting-for-learners/raw/main/README.md

3. Download a setup.py:

$ wget https://github.com/XXX/counting-for-learners/raw/main/setup.py

Modify nine shaded lines.

```python
import setuptools

with open("README.md", "r", encoding="utf-8") as fh:
    long_description = fh.read()

setuptools.setup(
    name="bugcount",
    version="0.0.10",
    author="XXX",
    author_email="XXX@keio.jp",
    description="A package for counting BLOB objects",
    long_description = long_description,
    long_description_content_type="text/markdown",
    url="https://github.com/XXX/counting-for-learners",
    project_urls={
        "Bug Tracker": "https://github.com/XXX/counting-for-learners",
    },
    classifiers=[
        "Programming Language :: Python :: 3",
        "License :: OSI Approved :: MIT License",
        "Operating System :: OS Independent",
    ],
```
4. The structure of directory (src) and files are as follows:

```
$ tree .
 ├── README.md
 │   └── setup.py
 │       └── src
 │           └── bugcount.py
```

5. Go to your directory with setup.py and README.md. Run the following commands. $ python setup.py install

```
$ python setup.py sdist bdist_wheel
$ rm dist/*.egg
```

6. Install twine and create a .pypirc file at your home directory. cat .pypirc shows the content of .pypirc file. password can be obtained from your pypi account settings: Click Two factor authentication (2FA) to generate API token. password starts with pypi-

```
$ cd
$ cat .pypirc
[pypi]
```
username = __token__

password = pypi-xxxxxxxx

$ pip install twine

7. The following command will automatically upload three files.

$ twine upload dist/*

CONCLUSION

The proposed program (bugcount) is one and only one open-source software without machine learning for counting aggregated objects. The bugcount application does not need time-consuming annotation tasks unlike existing machine learning methods. The bugcount program is available in public and can be installed by PyPI packaging for experimenting own counting disaggregated objects such as pests. PyPI packing allows bugcount to run on Windows, MacOS, and Linux operating systems with Python installed on your system. The proposed program lightly running on the conventional inexpensive CPU machines has a tunable Canny coefficient so that it can achieve perfect counting of disaggregated objects. The bugcount application has been downloaded 12344 times worldwide so that its usefulness, practicality and applicability was justified by the large number of downloads. However, as a limitation of the algorithm, the proposed method is not capable of classifying aggregated objects or overlapped objects. As future scope, a powerful disaggregation scheme is needed and should be integrated with the proposed method. This paper introduces generative AI as a tool for solving problems and addressing issues without relying on human experts in programming, installation, and debugging. Generative AI may be an indispensable assistant for many tasks, such as understanding technical terms and providing solutions for encountered problems.

Declarations

Acknowledgement: Not applicable

Disclosure of potential conflicts of interest: The author has no conflict of interest.

Clinical Trial Study: Not applicable

Dual Publication Statement: The work described has not been published previously and is not under consideration for publication elsewhere.

Data Availability Statement: Not applicable

Funding: This research has no fund.
Compliance with Ethical Standards

Research involving human participants and/or animals: Not applicable

Informed consent: Not applicable

References


15. Takefuji (2022) bugcount is a PyPI package for counting BLOB objects [Source Code]. https://doi.org/10.24433/CO.4823110.v1

Figures

(a) original picture  (b) gray picture  (c) blurred picture  (d) edged picture

Fig. 1 Four pictures (original, gray, blurred, and edged)

See image above for figure legend
**Fig. 2 Generated final result (Canny coefficient=75)**

Figure 2

See image above for figure legend

**Fig. 3 Generated final result (Canny coefficient=10)**

Figure 3
Fig. 4 Generated final result (Canny coefficient=100)

Figure 4

See image above for figure legend

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- APPENDIX.docx