

# Data Labeling: How AI Can Streamline Your Data Labelling

BYO ML-assisted labelling tool on Colab and try your hand at a game-changing no-code data labeling tool from Datature



Lai Woen Yon · [Follow](#)

Published in Towards Data Science · 10 min read · Dec 20, 2021



217



3



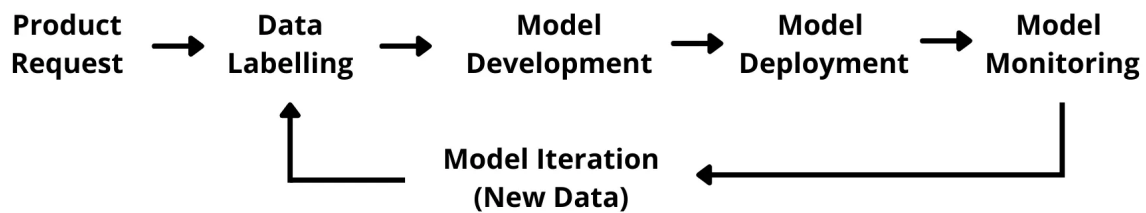
Photo by [daniyal ghanavati](#) from [Pexels](#). Image by Author.

Top 3 takeaways after reading this post:

1. How to create your own machine learning labelling tool using PixelLib on Colab.
2. Discover how Intellibrush can help you accomplish better labels, faster and without any code.
3. Key considerations when deciding on building vs buying an AI-enabled data-labelling tool

## **AI + Human Annotators = Fast Annotations Process**

The typical data science product development process is as follows:



Typical Product Development Process. Image by Author.

As part of the whole product pipeline, data labelling takes up most of the time. When it comes to data labelling, we engage human annotators to help label a large collection of unstructured data like images or text. Companies that are less concerned about privacy may outsource their labelling efforts to third-party labellers. However, in the event the labelled data involves sensitive data such as a customer's personal information or company IP, outsourcing is no longer an option and companies are faced with setting up their own data labelling team in-house which presents a whole new set of challenges.

Teams usually comprise multiple data labelers as well as a data engineer, where labelers are responsible for annotating data and ensuring that data is cleaned and ready to be ingested for model training. The data engineer, on the other hand, must be familiar with the end use-case of the machine learning application to provide a high overview of label consistency and benchmarks to be met by the labelers as well as ensure that biases are kept to a minimum. They may do so by setting up a labeling guide or performing consensus checking across different labelers to establish a baseline standard — because after all, garbage in is garbage out, and in this case, the performance of a machine learning model like an object detector heavily depends on the quality of labeled data.

But how can AI help label data especially for use cases such as drug research which requires a highly skilled and trained research scientist? This is where AI-enabled tools come into play where traditional statistical models are used in conjunction with pre-trained machine learning models to expedite the annotation process. An example of an AI-enabled labeling tool is IntelliBrush and as seen in the video below, a pixel-perfect mask annotation is done in a single click and is more than 10x faster compared to a regular non-AI enabled tool!



## The comparison of labelling efficiency with and without AL-assisted labelling

Lai Woen Yon

00:29

Photo by [daniyal ghanavati](#) from [Pexels](#). ML-assisted Tool (IntelligentBrush) by [Datature](#). Video by Author.

### Building Your Own ML-assisted Labelling Tool

Now that we have seen the capabilities of an AI-enabled tool, let's try and build our very own ML-assisted tool. To illustrate, I am going to use image segmentation as an example. The same concept can be applied to other machine learning tasks as well.

Medium

Search

Write

Sign up

Sign in



PixelLib to quickly segment objects in the images.

#### How we built an easy-to-use image segmentation tool with transfer learning

Label images, predict new images, and visualize the neural network, all in a single Jupyter notebook (and share it all...

[towardsdatascience.com](https://towardsdatascience.com)

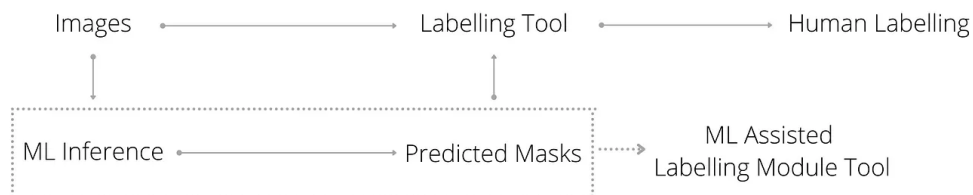
## Real Time Image Segmentation Using 5 Lines of Code

Perform accurate and fast object segmentation in images and videos with PixelLib

[towardsdatascience.com](https://towardsdatascience.com)

### Overview

Here is a summary of the whole pipeline. In the normal pipeline, human annotators can examine the images directly via a labelling interface. In order to make machine learning part of the labelling process, we must add a module called ML Assisted Labelling Module that enables user modifications of the machine-predicted labels directly on the labelling interface.



Simple Pipeline For Our ML-Assisted Labelling Tool. Image by Author.

### Image Segmentation Model

This demo will use PixelLib, a library for segmenting objects in images and videos. I chose PixelLib because it is easy to use and it provides rapid detection, which helps reduce time spent on the ML Inference.

Below is the code for using PixelLib for inference

```
# install pixellib
pip install pixellib

# download model pretrained weights
wget -N
'https://github.com/ayoolaofenwa/PixelLib/releases/download/0.2.0/p
ointrend_resnet50.pkl'

# instantiate model and load model weights
ins = instanceSegmentation()
ins.load_model("pointrend_resnet50.pkl", detection_speed='rapid')

# inference
result = ins.segmentImage(img_path, show_bboxes=False)
```

Masks can be extracted from the result and applied to your original image.



Before



After

Photo by [Pixabay](#) from [Pexels](#)

## Colab Demo

Here is the code for building an ML-assisted labelling tool on Colab

The following gif shows what the UI will look like once you run the code. You can label a bird with the lasso selector.

Gif by Author.

Alternatively, you can click the '**ml assisted**' button and the bird will be selected automatically by the machine learning model. Then, you can continue adding the missing pieces with the lasso selector tool.

Gif by Author.

You can use the demo to see how the ML-Assisted Tool can help you minimize your labelling efforts although this may not be very scalable if your team has multiple labelers.

Based on the demo, you can see how the ML-Assisted Tool can help you minimize your labelling efforts. However, further experimentation with the tool makes me conclude the following drawbacks.

Top highlight

1. Implementation is not scalable especially for internal teams due to each member having to load and save their raw image files and annotations separately.
2. PixelLib is trained using a Mask R-CNN model and the COCO dataset, thus custom objects that are unique to your use-case may not be detected accurately — defeating the purpose of having an in-house labeling team if the tool is unable to detect custom objects or object not commonly found in public datasets.
3. No shared access to dataset and labels — PixelLib may work if you are using it for a side project where only you, require access to the images and labels — however, that is often not the case in an organization where data engineers, labelers, and PM's work together to drive project success. This may pose as problems during the model iteration phase as debugging labels and images will be tedious no doubt.

You will notice that I have placed a high emphasis on the usability and collaborative-ness of the tool — as ML production is often a team effort and working in silos will often lead to delays. Read on to find out my top 3 considerations when looking for a data labeling tool!



## Quick Annotation Without Any Code

Some companies may not have the time or expertise to build their own labeling software, hence they look for off-the-shelf solutions.

I was drawn to a product developed by a company called **Datature**. Datature is a no-code MLOps platform that provides cloud-based workflows for data labeling and model training. The company recently launched a product called **IntelliBrush**, which is an AI-enabled data labeling tool that is designed to help companies boost their labeling productivity and efficiency so as to reduce the time taken to develop a fully working computer vision model. I have had the chance to try out their latest product and I am eager to share with you my own experience.

Feel free to sign up here if you cannot wait to try out this newest product:

<p><b>Datature IntelliBrush - Labelling Data 10x Faster with AI</b></p> <p>IntelliBrush is the fastest way for data labelling teams to label complex image data. Perfect for annotating datasets...</p> <p><a href="https://datature.io">datature.io</a></p>	
--	--

### What is IntelliBrush?

IntelliBrush is a built-in feature on Datature's Nexus platform. It uses machine learning models to predict the outline of your selected object. With this feature, users can quickly get pixel-perfect mask/bounding box annotations with just 1-2 clicks instead of needing to click multiple times on the border of an image if they were using a regular polygon tool, or trace the outline of an object where the margin of error tends to be high.

Furthermore, one thing I like about IntelliBrush is that it is continually tuned to make sure that it improves over time, and I can even select the level of granularity using Intelli-Settings which is great for when my image simply contains a single object or when it contains multiple smaller objects. Check it out on a variety of objects here:

Here are some highlights of IntelliBrush Adaptive Mode. Video by Author.

If you're a company or small startup looking for a platform to label and train a computer vision model using your custom data, you may wonder what factors you should consider before investing in an AI-assisted labelling tool. The following are 3 criteria I believe are important to consider and why I believe Datature's IntelliBrush is a great candidate to consider:

1. **Cost.** Whenever we develop a software system in-house, we must take into consideration the cost. An example of such a cost would be to hire a group of software engineers to develop the labelling tool, as well as a group of UX designers to design a user-friendly interface. Moreover, it might be necessary to hire a team of data scientists to build machine learning models if you wish to build a tool like IntelliBrush. Hence, when choosing a labelling platform, considering the cost of building, and maintaining the software is imperative. Datature's Nexus platform (with or without IntelliBrush) has plans catering to all kinds of businesses, whether you're just starting out with computer vision models or have a dedicated team of labelers looking for a platform to handle a high volume of data labelling tasks. They do have a free plan that comes with limited access to IntelliBrush which is great for teams who like to "try before they buy".

2. **Ease of use.** There are already ML-assisted solutions on the market, but some require you to develop your own machine learning models to support these features which is another challenge and can easily push your development timeline back by a couple of weeks and months. Another important fact to consider is that expert human annotators tend to be strapped for time and cannot afford to dedicate an entire day to labeling data — which is why an interface that is as streamlined as possible will go a long way in increasing efficiency.
3. **Flexibility.** Depending on the kind of models your team is developing, the labeling tool should also support both complex polygons and bounding boxes. In addition, the tool shouldn't be limited to common objects but the underlying algorithm should be able to detect new data that has never been seen before right out of the box. **This is a highlight of IntelliBrush as no pre-training is required, meaning it will work on any custom object as well!** As I mentioned above, the demo shows only how IntelliBrush segments images, but there are also several other tasks that can be supported, such as generating bounding boxes for object detection models.

### **How to use IntelliBrush?**

1. Log in to your [account](#), create a new project and upload your images
2. Open the web-based Annotator and create your first label.
3. Select IntelliBrush on the right-hand panel or use the hotkey “T”.  
(IntelliBrush should activate upon signing up. If it doesn't, you can apply [here](#) for early access.)
4. By left-clicking the center of the object of interest, the object will be masked immediately.
5. The mask can be edited by using a right-click to denote regions that are ‘out-of-interest’ if you are not satisfied with the generated mask and can be refined as many times as you need.
6. Once you are happy with the mask, you may press space to commit the label.

Check out the video below to see how it works in action. I have also tried out different images with the tool!

An example of how to use IntelliBrush for various scenarios. Video by Author.

## Conclusion

If your team is constantly held back by the lack of high-quality labeled data, perhaps an ML-assisted labeling tool can help to increase your team's productivity. If you're looking for a fast and accurate 'off the shelf' tool, IntelliBrush is an ideal candidate as it requires no prior model training and works even for never-before-seen images. Moreover, the company is actively improving the tool and plans to continue doing so with planned releases for QA checks on top of their existing collaborative labeling features. Finally, if you are interested in building your own computer vision model, here are some videos from [Datature](#) to help you kickstart your own computer vision project using the Nexus platform — all without code.



## About the Author

Woen Yon is a Data Scientist based in Singapore. His experience includes developing advanced artificial intelligence products for several multinational enterprises.

Woen Yon works with a handful of smart people to offer web solutions including web crawling services and website development for local and international start-up business owners. They are well aware of the challenges of building quality software. Please do not hesitate to drop him an email at [wushulai@live.com](mailto:wushulai@live.com) if you need assistance.

He loves making friends! Feel free to connect with him on [LinkedIn](#) and [Medium](#)

### Lai Woen Yon - Medium

Read writing from Lai Woen Yon on Medium. Data Scientist, TDS contributing writer. I love making friends from all...

[laiwoenyon.medium.com](https://laiwoenyon.medium.com)

Artificial Intelligence

Data Science

Computer Vision

Startup

Machine Learning



## Written by Lai Woen Yon

213 Followers · Writer for Towards Data Science

Follow



Data Scientist, TDS contributing writer. I love making friends from all around the world!

---