

Possible use cases of the vision setup

Setup:

1. Camera
2. Tensor Processing device(Jetson Nano/Intel Neural Stick/etc.)
3. Microcontroller

Scenario:

Though the idea can be applied to various scenarios. I have considered the case where the subject is hiking as was suggested.

Applications:

This setup can benefit the subject in following two ways:

1. Classification of local objects

When going in the wild, a wide range of vegetation can be seen, some plants are edible, others are poisonous. And some we straight away find interesting and are curious to know more about them. On an average around 60 people die in the US in a year from eating poisonous plants and hundreds end up in hospitals.

The idea here is, when a certain person is hiking with the setup. The camera can extract frames from the video and we can process those frames in realtime to get information about the plants. We can then notify the user of certain properties of those plants/mushrooms.

This can be a great device for hikers.

2. Identification of the surface subject walks on

To keep track of the amount calories burned, people rely on fitness bands, which calculates it by using features like number of steps and heart rate.

To increase the efficiency of the process we can use a small neural network to estimate the surface and calculate the approximate slope. This

can also help user get a better track of how much he/she climbed uphill and downhill.

Execution:

Application 1 can be executed in two ways. Either building our own deep learning model or using third part services like [google lens](#) or [plant id](#). I used google vision API, results were not so good, however plant.id give great results.

Here the frames would be sent and results would be retrieved using the API, hence no computation needed.

But in activities like hiking, it would be difficult to connect to these remote servers. Better option would be to use micro GPUs like Jetson Nano, and train our own models on available [datasets](#).

I would suggest an object detection network first to localize plants in the frames, another network classifying the part of the plant(root, flower, leaf, etc.), and a final network classifying the plant. Although with more experimentation, can come up with a better architecture.

In case of Application 2, even though there are different methods to approximate depth of objects using CNNs, a more accurate approach would be using depth camera instead of a 2d camera. We can map the surface from the depth camera and real time and calculate the slope with better accuracy, helping the subject to track the activities in a better way.