

**Image Scraping and Classification Project**

Submitted by:

MAHESH AGASHIMANI

**ACKNOWLEDGMENT**

I take great pleasure to thank and acknowledgement the allowance by Data Trained Education and permission by Flip Robo. I extend whole hearted thanks to them I worked and learned a lot and sharing me the knowledge and experience.

Data Trained Education and Flip Robo provided training is the very important to completion of project.

**INTRODUCTION**

* Business Problem Framing

Images are one of the major sources of data in the field of data science and AI. This field is making appropriate use of information that can be gathered through images by examining its features and details. We are trying to give you an exposure of how an end to end project is developed in this field.

* Conceptual Background of the Domain Problem

The idea behind this project is to build a deep learning-based Image Classification model on images that will be scraped from e-commerce portal. This is done to make the model more and more robust.

* Review of Literature

If you have ever tried performing image recognition using deep learning, you know the importance of a good dataset for training. However, finding sufficient images for training is not always easy, and the accuracy of your model is directly dependent of the quality of the training data.

* Motivation for the Problem Undertaken

There are techniques you can use to supplement the images dataset that they use for training. One of the techniques is called *image data augmentation*.

**Analytical Problem Framing**

* Mathematical/ Analytical Modeling of the Problem

Image data augmentation is a technique that creates new images from existing ones. To do that, you make some small changes to them, such as adjusting the brightness of the image, or rotating the image, or shifting the subject in the image horizontally or vertically.

Image augmentation techniques allow you to artificially increase the size of your training set, thereby providing much more data to your model for training. This allows you to improve the accuracy of your model by enhancing the ability of your model to recognize new variants of your training data.

* Data Sources and their formats

I will first demonstrate the various image augmentation techniques using Python and Keras. If you want to try along, make sure you have the following software and packages installed:

**TensorFlow**. You can install TensorFlow using the command

!pip install tensorflow

Once selenium and tensorflow are installed, create a new Jupyter Notebook.

Data Collection Phase: In this section, you need to scrape images from e-commerce portal, Amazon.com. The clothing categories used for scraping will be:

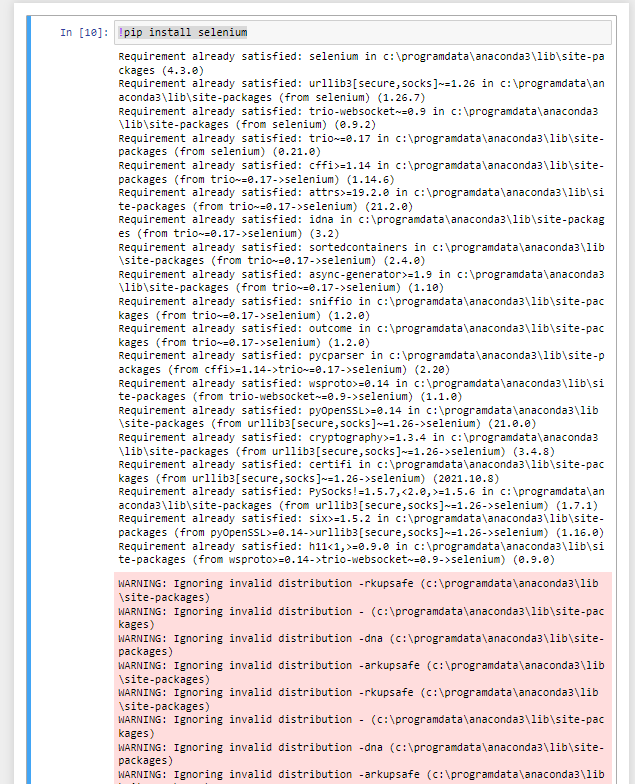
* Sarees (women)
* Trousers (men)
* Jeans (men)

You need to scrape images of these 3 categories and build your data from it. That data will be provided as an input to your deep learning problem. You need to scrape minimum 200 images of each categories. There is no maximum limit to the data collection. You are free to apply image augmentation techniques to increase the size of your data but make sure the quality of data is not compromised.

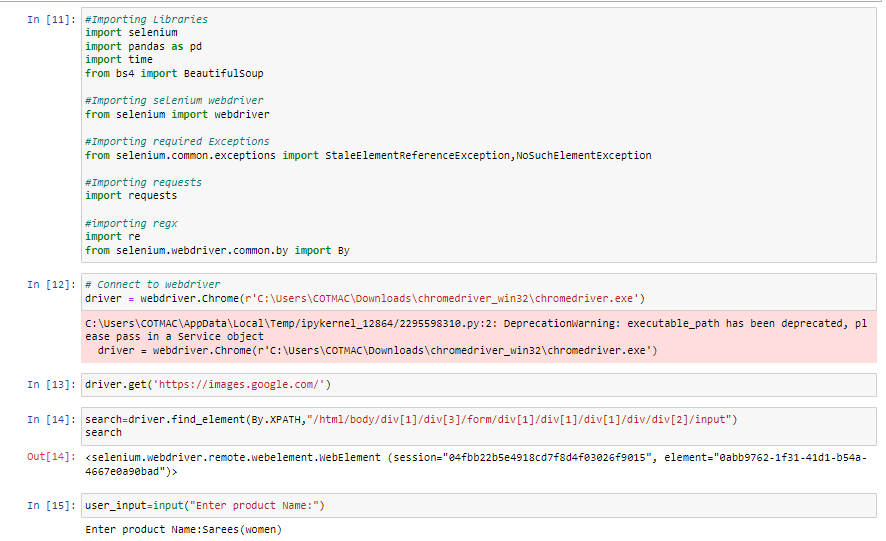
Importing the Libraries:

!pip install selenium

Downloading the Sarees (women) images in the given path using webscrapping -



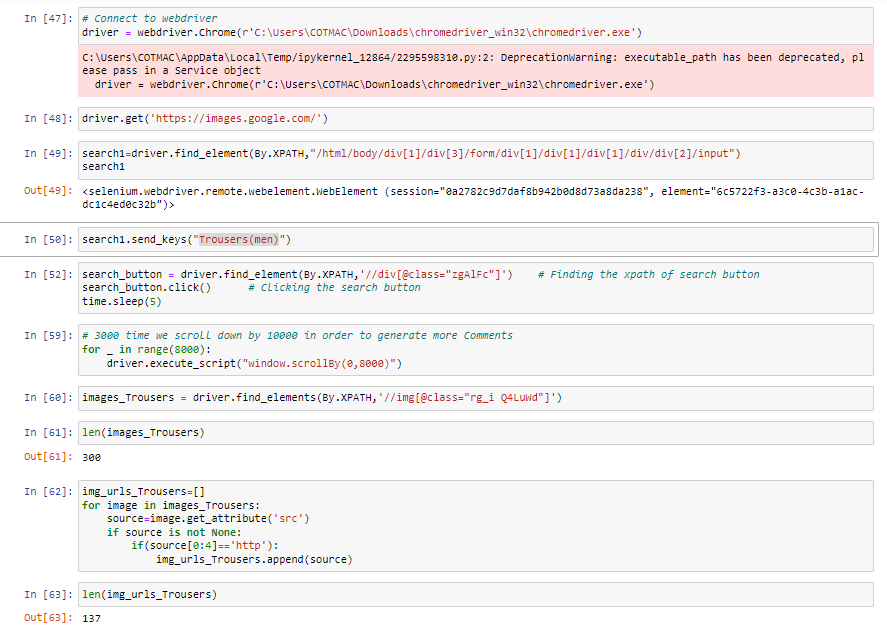
. Loading the Dataset:







Downloading the Trousers(men) images in the given path using webscrapping –





Downloading the Jeans(men) images in the given path using webscrapping -





* State the set of assumptions (if any) related to the problem under consideration

As given in datasets my assumption is predicting image scrapping Classification Problem.

* Hardware and Software Requirements and Tools Used

The needed time to train the model depends on the capability of the used system during the experiment. Some libraries use GPU resources over the CPU to take a shorter time to train a model.

|  |  |
| --- | --- |
| Operating System | Windows 10 |
| Processor | Core i7 |
| RAM | 16GB |
| Graphics card | 1080 TI OC |

Also we are using Jupiter notebook for running the code.

**Model/s Development and Evaluation**

Importing Libraries

! pip install tensorflow

!pip install image

*The ImageDataGenerator class allows your model to receive new variations of the images at each epoch. But do remember that it only returns the transformed images and does not add it to the set of images that you have.*

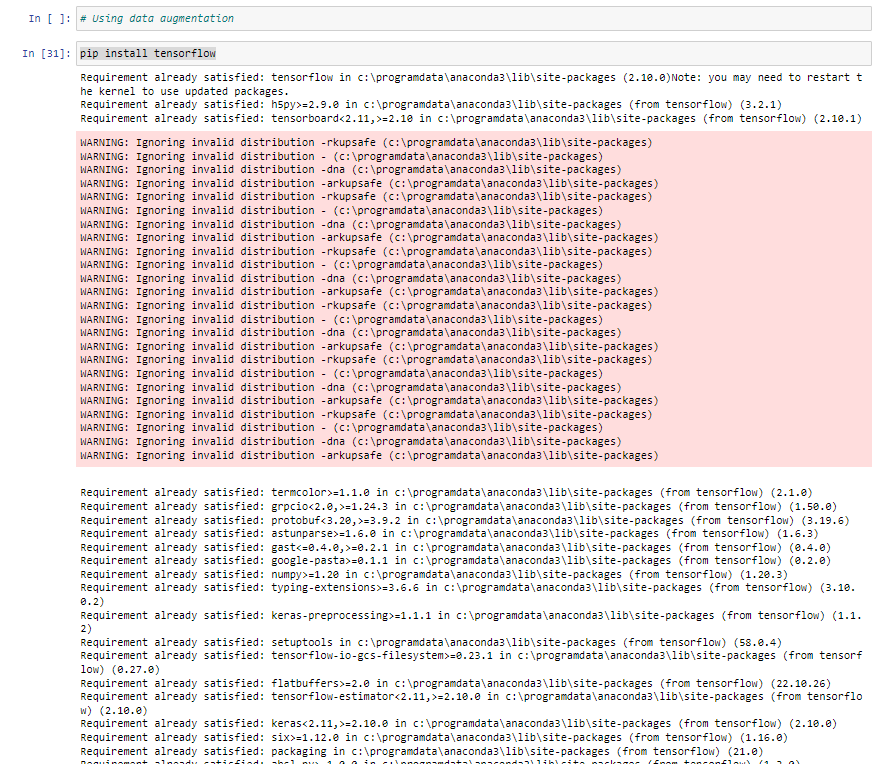
**Using Data Augmentation –**

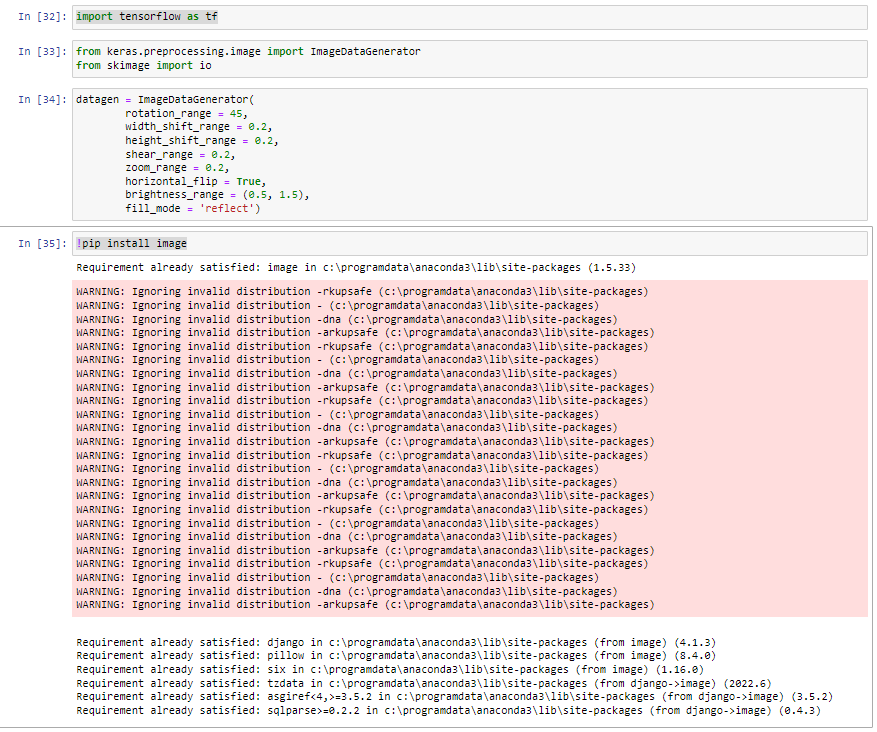
Image data augmentation is a technique that creates new images from existing ones. To do that, you make some small changes to them, such as adjusting the brightness of the image, or rotating the image, or shifting the subject in the image horizontally or vertically.

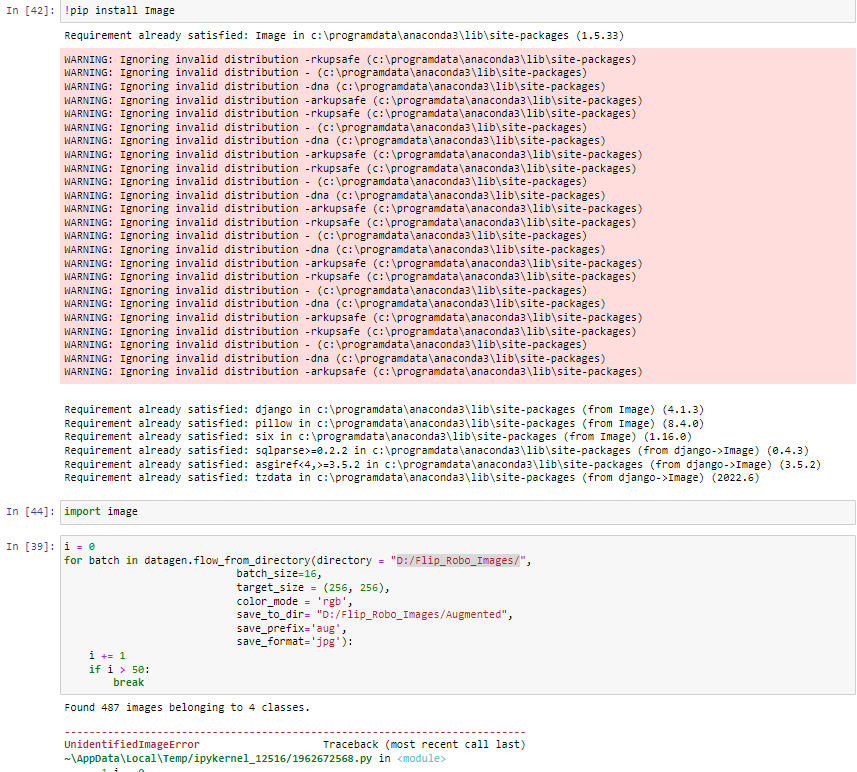
**Types of Image Data Augmentation**

Image augmentation comes in many forms, here are some of the common ones — Vertical shift, Horizontal shift, Vertical flip, Horizontal flip, Rotation, Brightness adjustment, and Zoom In/Out.

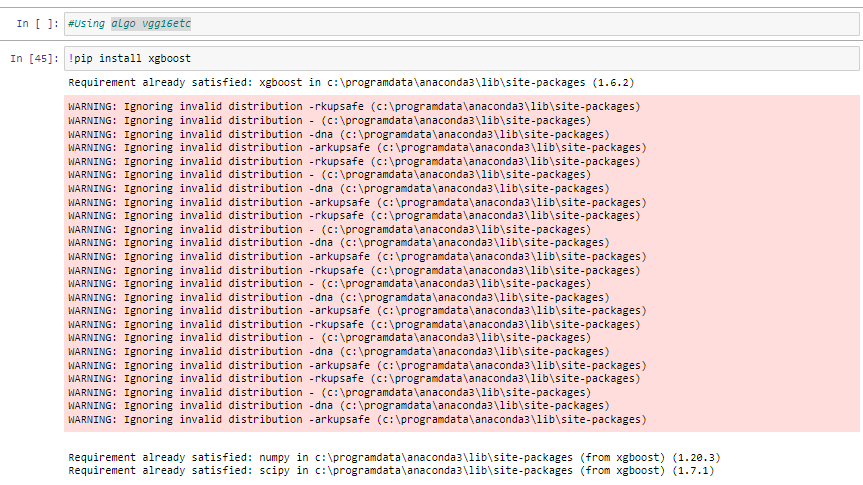
**Using data augmentation –**

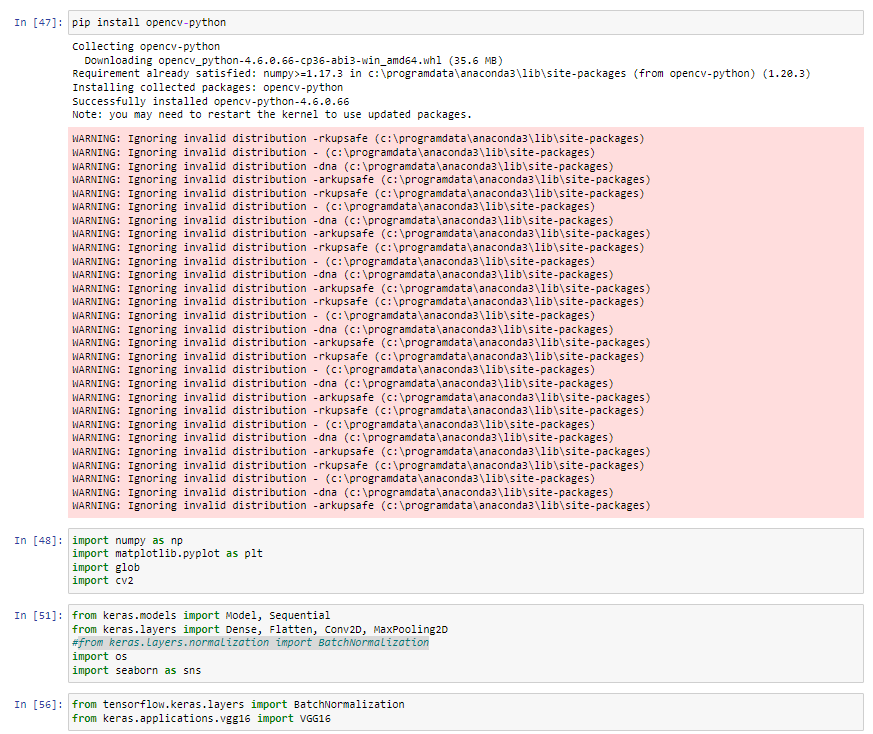


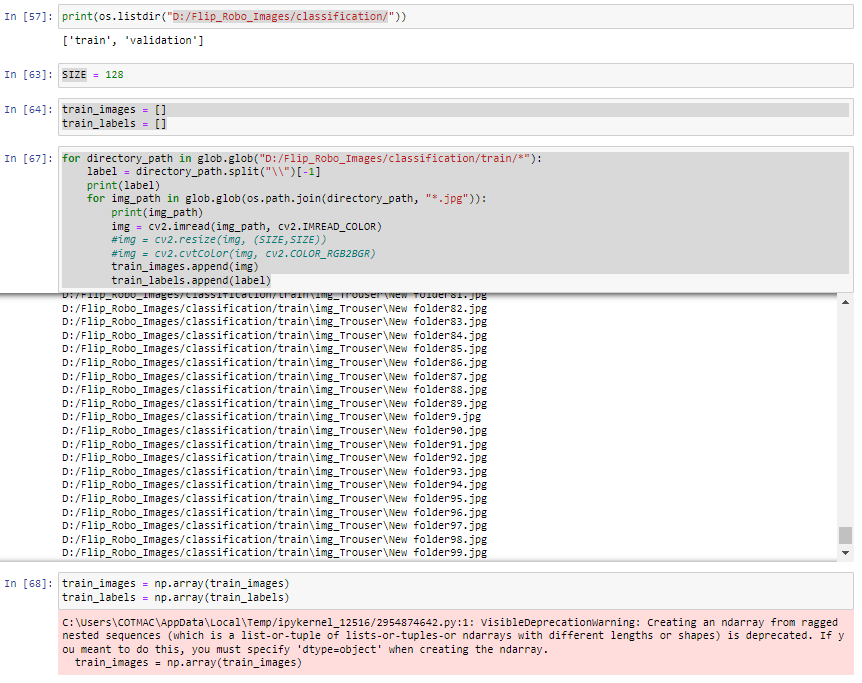


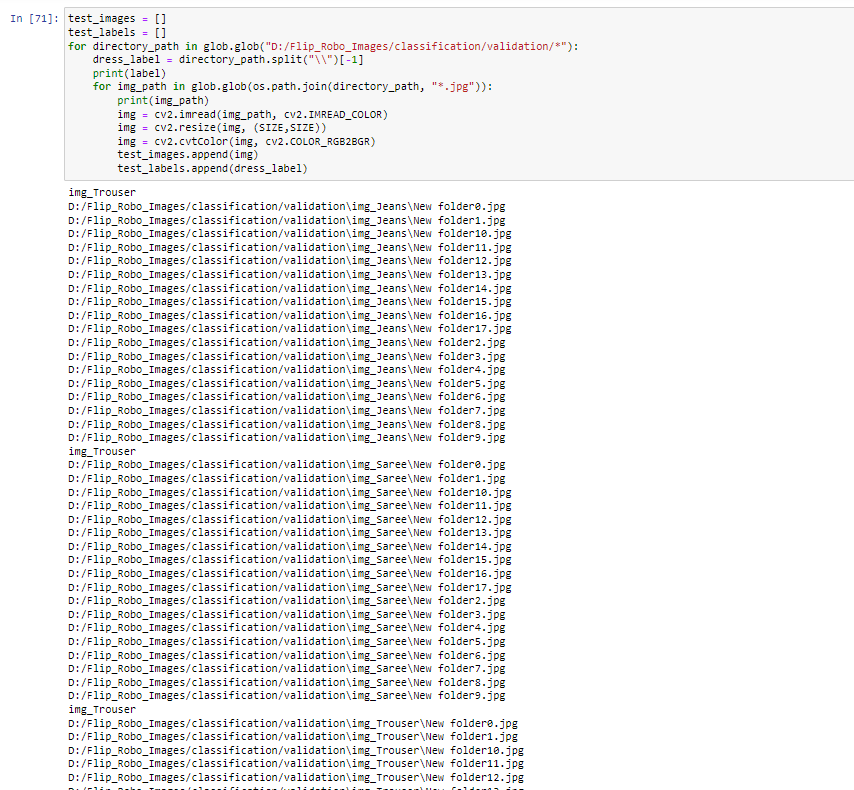


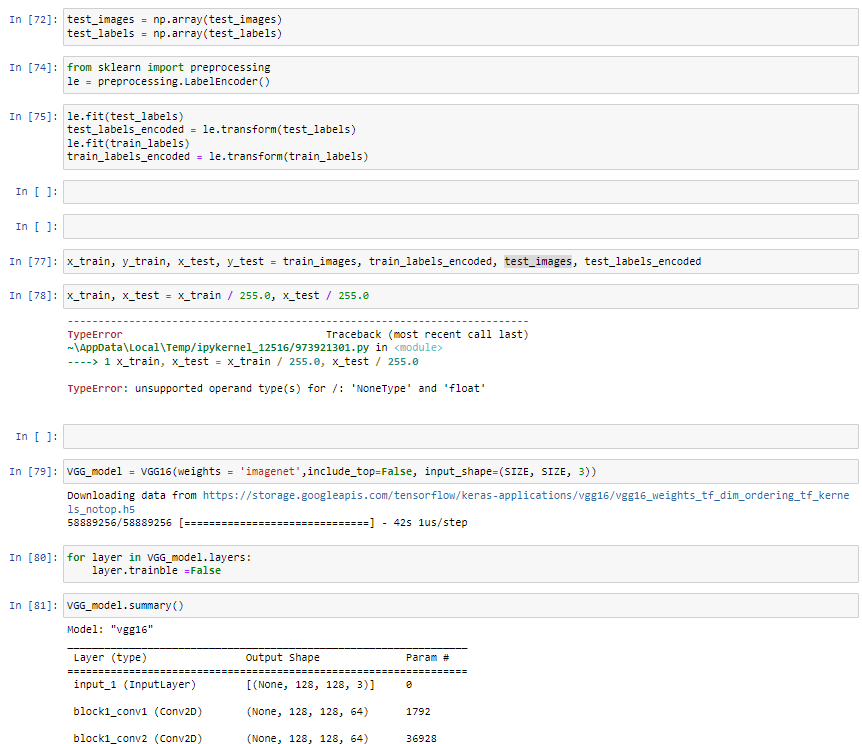
**Using algo vgg16etc –**











**CONCLUSION**

Data Augmentation is used to Image editing like Vertical Shift, Horizontal Shift, Horizontal Flip, Rotation, Brightness.