

Driver License YOLOv3 Dataset Documentation

In this document, we report the description and construction of the dataset of images and labels used in order to train a YOLOv3 deep learning architecture to detect the presence of driver licenses of any United States state in an emergency medical response environment. The final weights obtained after training for 5500 epochs on this dataset can be found as **yolov3_training_last.weights**.

The overall unseparated dataset of all images is compressed within the **images.zip** file. Separate and categorized images are stored within the separately labeled folders. To preserve Google Drive storage, there are only a sample of representative images within each of the folders. Below, we describe the contents of each folder and how they were obtained.

NOTE: The file architecture is important for the python files to function. If you wish to adjust the file architecture, make sure to adjust the files accessed by adjusting the parameters at the top of the files.

Python Files: Within the folder, there are a few python files that were used to process labels and images or augment data.

- **annotate_negatives.py** creates blank labels for all .jpg images in the folder **no license**.
- **midvprocess.py** turns the .tif files into .jpg files in the midv500 file, and translates the boundary coordinates from [top-left corner coord, top-right corner coord, bottom-left corner coord, bottom-right corner coord] to [x center, y center, x relative width, y relative width], as required by YOLOv3.
- **data_augment.py** applies a random perspective transform to the images in the folder **flat/#** and pastes them onto background images in the folder **background**. It may also distort brightness and color for better training.

Online Datasets Used

- **MIDV500 - A dataset for identity documents analysis and recognition on mobile devices in video stream**

By Arlazarov, Bulatov, Chernov, and Arlazarov, the MIDV500 dataset contains iPhone and Android recordings of identity documents, such as driver licenses and passports, of a variety of different countries. Overall, there are 50 different documents, 5 different settings, and two different modes of recording, resulting in a total of 500 recordings. The frames of the first 3 seconds of these recordings are stored, at 10FPS, to produce images, resulting in 30 images per recording. The boundary box of the documents in these images are labeled in a JSON file. Overall, due to the fairly natural environment of these licenses, they represent good training data and will be used in the positive training set.

License: Creative Commons Attribution-ShareAlike 2.5 Generic License

Source: V. V. Arlazarov, K. Bulatov, T. Chernov and V. L. Arlazarov, “A dataset for identity documents analysis and recognition on mobile devices in video stream”, *Comput. Opt.*, vol. 43, no. 5, pp. 818-824, 2019.

- **Open Image Dataset - An online open source dataset containing images that are pre-labeled to contain certain objects.**

By Kuznetsova et. al in 2020, the Open Image Dataset “is a dataset of ~9M images annotated with image-level labels, object bounding boxes, object segmentation masks, visual relationships, and localized narratives.” Using the `OIDv4_ToolKit` (https://github.com/EscVM/OIDv4_ToolKit), we obtained the images and labels in available classes that we figure the algorithm could falsely label as a license. In particular, the images and labels belonging to the categories “Accordion”, “Bathroom cabinet”, “Billboard”, “Book”, “Briefcase”, “Calculator”, “Door”, “Mobile phone”, “Paper towel”, “Refrigerator”, “Suitcase”, and “Table” were obtained as a negative set.

License: CC by 4.0 International

Source: Kuznetsova, Alina, Hassan Rom, Neil Alldrin, Jasper Uijlings, Ivan Krasin, Jordi Pont-Tuset, Shahab Kamali, et al. 2020. “The Open Images Dataset V4.” *International Journal of Computer Vision* 128 (7): 1956–81. doi:10.1007/s11263-020-01316-z.

- **Stanford Background Dataset - A lightweight scene understanding dataset**

Originally by Gould, Fulton, and Koller, the Stanford Background Dataset contains images with labels that segment and label each pixel in the image. The images contain outdoor scenes, are approximately 320-by-240 pixels, contain at least one foreground object, and have the horizon position within the image. Because of the criteria for these images, we believe that they are representative of an environment wherein the YOLOv3 weights will be used. These images will both serve as control images, and as background images for data augmentation.

License: Creative Commons Attribution-ShareAlike 2.5 Generic License

Source: Gould, Stephen, Richard Fulton, and Daphne Koller. 2009. “Decomposing a Scene into Geometric and Semantically Consistent Regions.” *2009 IEEE 12th International Conference on Computer Vision*. doi:10.1109/iccv.2009.5459211.

- **“What a Driver’s License Looks like in Every State” - Flat license images**

A collection of images of driver licenses that are viewed head-on with no rotation. They will be used in data-augmentation to create unique images for training.

License: GLU

Source: Shaw, Gabbi, and Frank Olito. 2020. "What a Driver's License Looks like in Every State." *Insider*. Insider. January 21.
<https://www.insider.com/what-drivers-license-looks-like-in-every-state>.

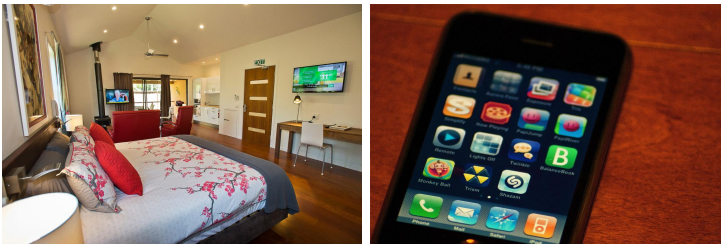
Categories of Images:

Category	Number of Images w/ Labels
No License	7209
Prev Yes License	22
New Yes License	11700
Custom Yes License	23834

- **No License:**

These images were obtained from the Stanford Background Dataset and the Open Image Dataset and serve as negative training for the YOLOv3 algorithm. There are outdoor images, and images containing objects that we determined that the algorithm may confuse as licenses. Each image's label is empty.

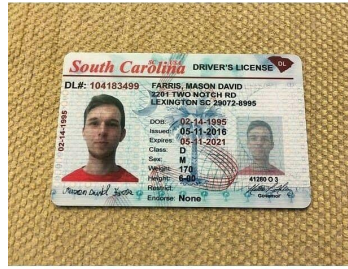
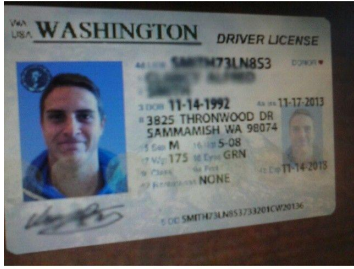
Examples:



- **Prev Yes License:**

Grabbed off the internet from a cursory Google search of 'drivers license', these images contain relatively natural settings for training.

Examples:



- **New Yes License:**

These images were obtained by processing the midv500 dataset. The `midvprocess.py` turns the .tif files in the `midv500/ground_truth` files into .jpg files, and translates the boundary coordinates from [top-left corner coord, top-right corner coord, bottom-left corner coord, bottom-right corner coord] to [x center, y center, x relative width, y relative width], as required by YOLOv3. The new images and labels are stored in **new yes license**.

Examples:



- **Custom Yes License:**

These images were made by pasting images from `flat/#` on top of those in `background`, which respectively contain flat images of driver licenses from the article “What a Driver’s License Looks like in Every State” and the representative background images from the Stanford Background Dataset. The `data_augment.py` file applies a random perspective transform to the images in the folder `flat/#` and pastes them onto background images in the folder `background`. It may also distort brightness and color for better training.

Examples:

