Problem Statement: Implement semantic segmentation using a deep convolutional network on a drone footage of different building surfaces.

Objective: To segment the parts of a building so that we can tell a drone where to look for stains.

Improvements: We can run our model through a 3D map so that we can eliminate human interaction to create a fully automated system. Adding more labels to the dataset to provide more sophisticated object detection using advanced Neural Network algorithms.

Motivation: The major pain point of the traditional approach is safety. One thing we would like to mention in this write up is how Lucid was created. During our discussion, we asked the board members “What inspired you to come up with this idea?”, to which they replied by mentioning a tragic incident that cost a person their life during one cleaning assignment. This inspired us into believing in the idea even more and develop a solution with 100% safety.

Tools Used:

DeepLap: DeepLab is a state-of-art deep learning model for semantic image segmentation, where the goal is to assign semantic labels (e.g., person, dog, cat and so on) to every pixel in the input image. Current implementation includes the following features:

CNN: We used the pretrained weights from the Deeplab checkpoint Xception-65

TensorFlow: TensorFlow™ is an open source software library for high performance numerical computation

LabelMe:Labelme is a graphical image annotation tool inspired by http://labelme.csail.mit.edu. It is written in Python and uses Qt for its graphical interface.

Approach:

We have labeled 50 images which are generated from drone video. The tool we used for labelling the data is LabelMe. This will have labels brick, window, casing and

background. We then use Labelme to annotate our image data. Labelme is an open source tool that allows us to label pixels in an image and writes the annotation data to a json file. This file contains all relevant image annotation data for example label, bounding box pixel list, encoded byte image data etc.
We need to serialize this data to a tensor so that the computation is faster in Deeplab. Wr converted our annotation data to a Tfrecord file format. It is optimized for use in Tensorflow in many ways. It makes it easy to combine multiple datasets and integrates seamlessly with the data import and preprocessing functionality provided by the Tensorflow library.

We use this data and pass it as a tensor to deeplab. We used various flags provided by Deeplabs built-in methods to fine-tune our model. We ran this on an NVIDIA GTX-1060 graphics card. This process took us about 40 minutes for 2000 iterations.

Challenges faced:

Firstly, we trained Xception-65 for classifying the different types of the surface but this is a classification algorithm and was unable to detect the windows. So, we passed another set of labeled images using YOLO Annotation Tool. We trained on a set of labelled images on YOLO v3 which is a state-of-art real-time object detection system.

Coming back to Deeplab, we realized that our data was class imbalanced, i.e. our major class background as brick pixels was way unproportionate as compared to other class label pixels. To fix this we tried adding more weights to the classes where the label pixels were outnumbered. This resulted in a slightly better result.

Results and Visualizations:

References:
[1]. Semantic Image Segmentation with Deep Convolutional Nets and Fully Connected CRFs Liang-Chieh Chen+, George Papandreou+, Iasonas Kokkinos, Kevin Murphy, Alan L. Yuille In ICLR, 2015.

[2]. B. C. Russell, A. Torralba, K. P. Murphy, W. T. Freeman. LabelMe: a Database and Web-based Tool for Image Annotation. International Journal of Computer Vision, 77(1-3):157-173, 2008

[3]. "Semantic Image Segmentation with Deep Convolutional Nets and Fully Connected CRFs"

[4]. Rethinking Atrous Convolution for Semantic Image Segmentation Liang-Chieh Chen, George Papandreou, Florian Schroff, Hartwig Adam. arXiv: 1706.05587, 2017

[5]. Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation Liang-Chieh Chen, Yukun Zhu, George Papandreou, Florian Schroff, Hartwig Adam. [link]. In ECCV, 2018

[6]. ParseNet: Looking Wider to See Better Wei Liu, Andrew Rabinovich, Alexander C Berg arXiv:1506.04579, 2015.