**Part 1 – Data Acquisition**

* Difficult to acquire such large amounts of data
  + Sources like shutterstock is expensive
* Possible image source: other open source machine learning database ([Dataturks](https://dataturks.com/projects/sheerun/rooms))
* Expand image pool using The Keras deep learning neural network library to artificially tweak images
  + Image data augmentation via the *ImageDataGenerator* class

**Part 2 – Data Labelling**

* The next problem lies in creating “correct” answers to teach the machine
* Using Amazon Mechanical Turk, the only available way to label images is by:
  + Categorising – identifying what the picture is
  + Isolating features – using a box to highlight a certain part of the image (this is the option that is most suitable in this case)
* This means that it will not feasible to highlight every single room edge in the image
* An alternative will be to box out every single visible corner in the image and train the machine to identify corners
* Then, an edge detection software can be run on each mini image of corners to get the edges
* The rest of the edges can be filled in with interpolation and regression
* Problem: how big is each box, and what sort of guidelines should be given to workers?