ODR Research: <https://medium.com/synapse-medicine/recognizing-drugs-on-medical-prescriptions-557b4e8103b2>

Why not use an existing OCR SDK:

* data protection
* commercial OCR SDKs are train to recognize text on many different kinds of images in everyday’s life. Train an model with specific case seemed more interesting and relevant

Our approach:

* Generate own custom dataset:
  + Using the English/Vietnamese public databases as reference vocabulary to generate an dataset. Because we want to familiarize our OCR system with drugs names
  + They  used 150x150 image’s size. Each of this contains one or two words randomly chosen in our reference vocabulary
  + Our goal are to generate pictures looking as realistic as possible: background color, font style, colors, text rotations, shadows, blurs, etc…
  + All this process have to chose randomly
  + They generate 12 milion pictures using this process. They also included empty images, colored background, random shapes such as lines, barcode etc
* CNN (InceptionV3), LSTMs and CTC loss
  + For using this approach, we can loading pre-trained InceptionV3 weight, we considerably enhanced the performance of the network.