1. 因為sigmoid的output範圍是(0,1)，在backpropagation的時候，梯度容易因為太多層使得其趨近於0，導致梯度消失，而ReLU的梯度都是1或0，較容易可以減少這種風險。
2. 一張含有 文字, 字型, 螢幕擷取畫面, 白色 的圖片

   自動產生的描述一張含有 文字, 字型, 行, 白色 的圖片

   自動產生的描述

一張含有 螢幕擷取畫面, 文字, 字型 的圖片

自動產生的描述

因為我不擅長通靈，所以試幾次後，最後我決定參考了網路上的方法。

在選值的過程中，我發現如果pool\_size選很小，跑超慢，可能是因為減少的feature就會比較少所以變超久；filter\_size則是越大跑越慢，我猜測是因為計算量的問題；然後layer太多，learning太小，速度直接慢到起床了還沒train好。

NN 一張含有 文字, 螢幕擷取畫面, 字型 的圖片

自動產生的描述model parameters -> 784\*128 + 128\*64 + 64\*32 + 32\*1 = 110624

CNN model parameters -> 3\*3\*1\*8 + 8\*8\*8\*1 = 584

Compare -> 110624 – 584 = 110040 ????