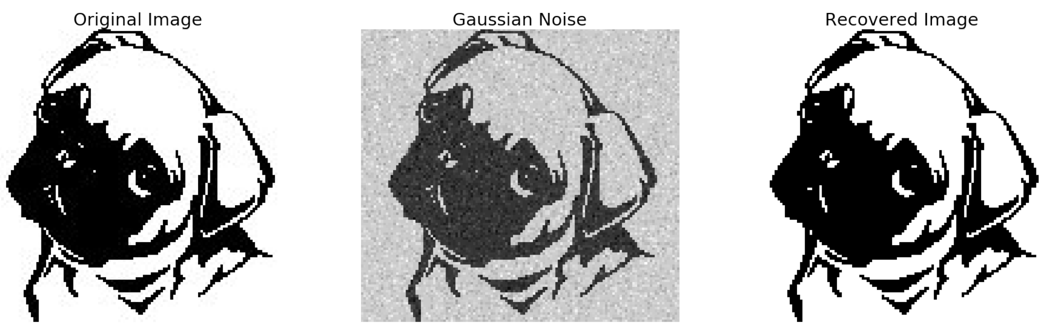
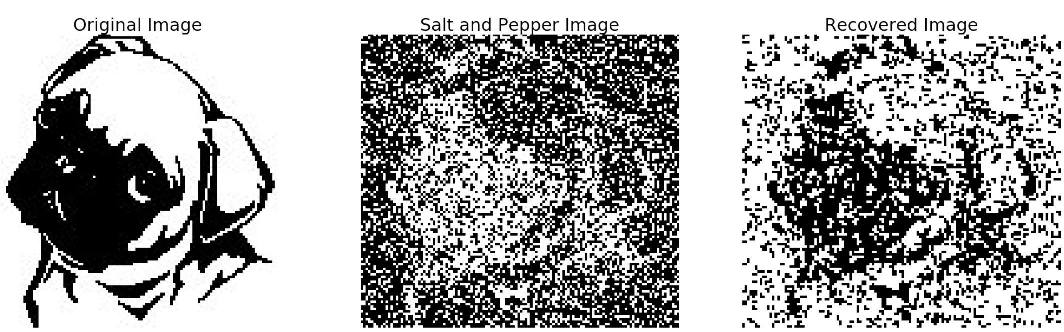
Inference

November 30, 2017

2

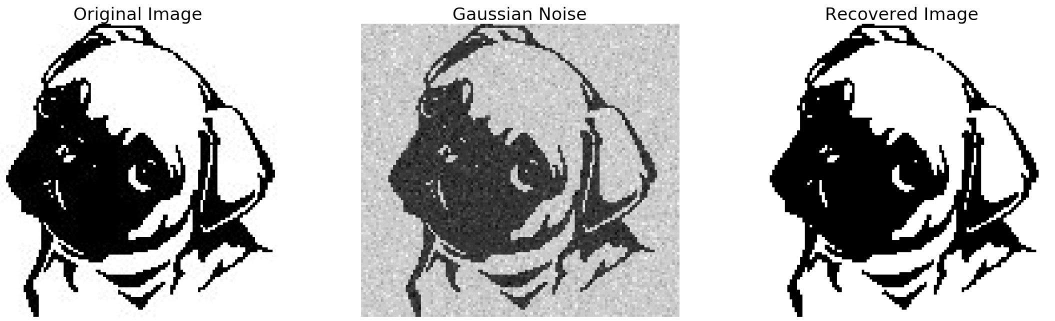


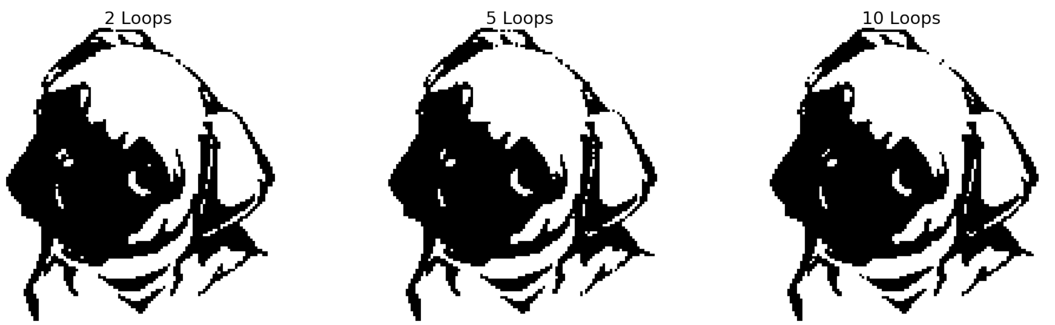
4

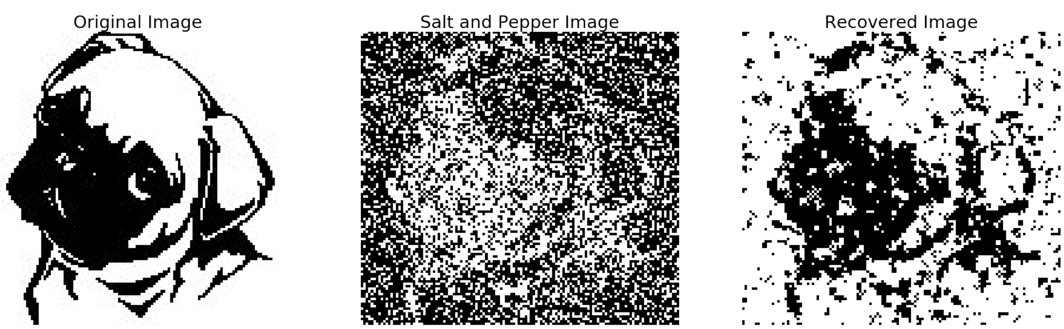


It takes 2 and 5 iterations respectively. For the salt and pepper image, this clearly still isn't a "decent" image, but for the gaussian, the image is very close to the original.

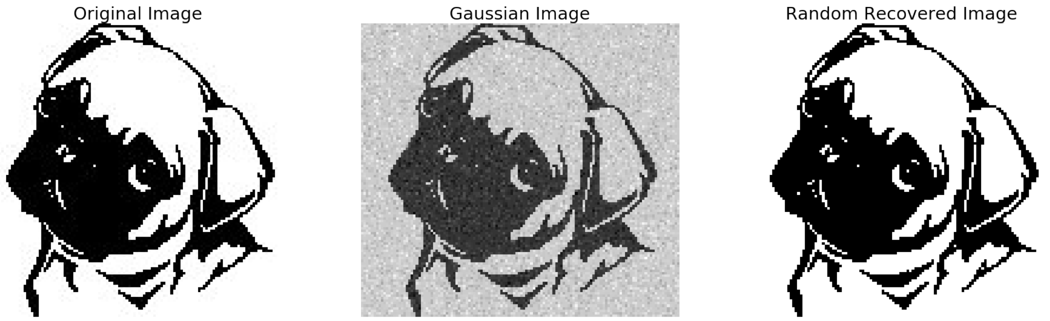
Q2



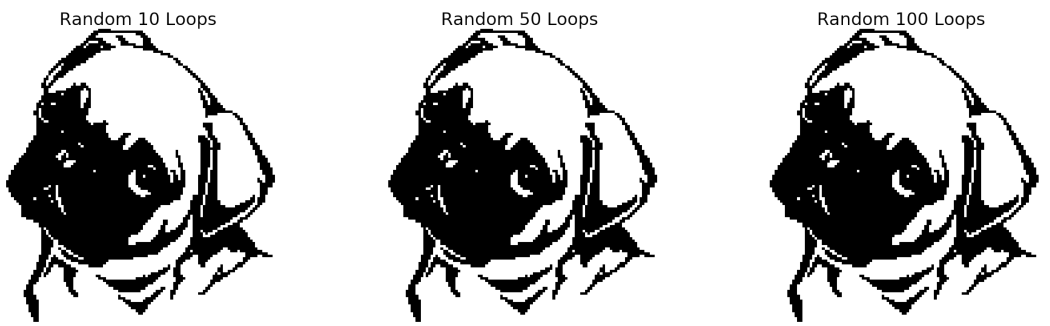




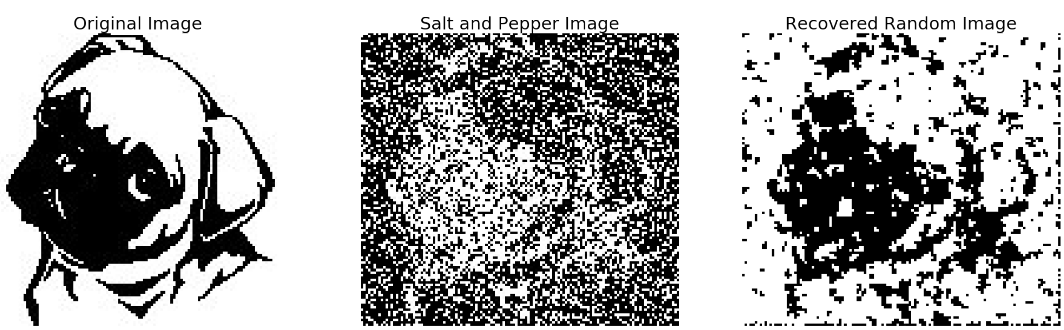

```
Number of iterations:  1
```




```
Number of iterations:  10
Number of iterations:  50
Number of iterations:  100
```




```
Number of iterations:  500000
```



3

When looping randomly, the gaussian noise image seems to near enough recover the image perfectly in what appears to be one iteration, whilst the S&P image seems to peak in quality at around the 500,000 iteration points. In this case, iterations are how many pixels are being updated.

4

As stated above, the image does not always improve, particularly with the salt and pepper image, after an optimal number of iterations, the image declines back into a less-recognisable mess. The gaussian example shows detrimental effects from looping more than a nice amount of times, as the image quality decreases the more loops are applied.

5

The difference between the two different forms is that in a case where our $p(x)$ is very unlikely, such as, using a very unfair coin, yet our approximation is $q(x)$, a normal, fair coin, our $q(x)/p(x)$ will produce a very large output (a large entropy) and therefore need lots of bits to model. On the other hand, with $p(x)/q(x)$, you get a significantly smaller number, therefore a small entropy. As an example, the two representations differ as in one, we are seeing how suitable $p(x)$ is to represent $q(x)$, whilst vice versa in the other example.

To reason about the difference between $KL(q(x)||p(x))$ and $KL(p(x)||q(x))$, we'll consider the specific case where we fix $p(x) = 0, q(x) > 0$. Using the definition of the KL Divergence, $KL(q(x)||p(x)) = \int q(x)log\frac{p(x)}{q(x)}$ we can compute that

$$lim_{p(x)\to 0}\, q(x)log\frac{p(x)}{q(x)} = -\infty$$

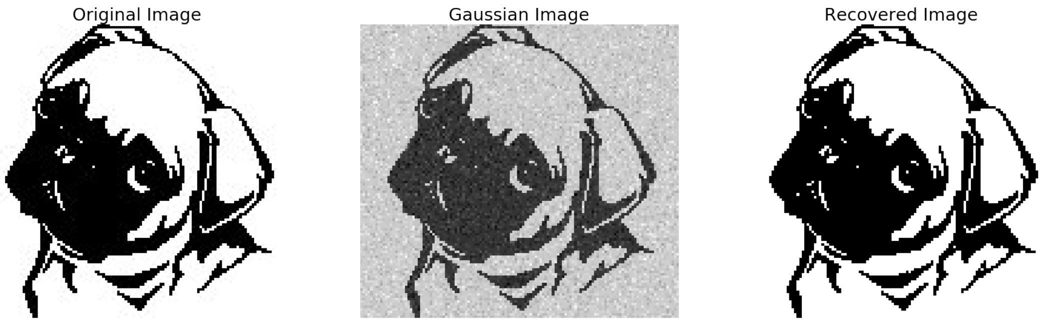
This would mean that if there is some region where we assign probability mass in $q(x)$ that $p(x)$ assigns no probability mass to, then the KL divergence will be of magnitude ∞.

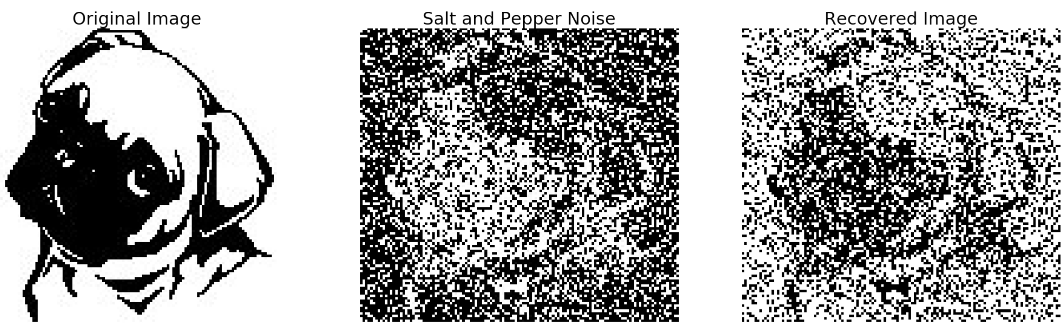
On the otherhand, if we fix $p(x), q(x)$ as before, and calculate $KL(p(x)||q(x))$ we get a very different result.

$$lim_{p(x)\to 0}\, p(x)log\frac{q(x)}{p(x)} = 0$$

What this suggests is that if there is probability mass at q(x), but none at p(x), it doesn't make the two any less similar. NEEDS FININSHING (WILL DO TOMORROW) https://link.springer.com/content/pdf/10.1007%2F978-3-642-00659-3.pdf <- helpful for this, pg 57

6





7

The variational bayes method, as with the others, produces a good quality retrieval of the gaussian image, but unlike the others, produces a fairly cleaner S&P image.


Original Image
Foreground
Background