**Cutout**

1. CNNs often susceptible to overfitting, cut out square regions to improve the robustness and overall performance
   1. Nice because compatible and easy to implement
2. Tens to hundreds of millions of params in modern networks - this is what can lead to overfitting
3. Drop out puts noise into neurons - this guy applies noise onto the images themselves
   1. **This is like drop out, but units are dropped at input layer of CNN**
   2. **And the dropouts are contiguous regions, not single pixels**
4. **Encourages network to better utilize full context of image, instead of the presence of a small set of specific visual features**
5. Very related to the ideas of data augmentation and dropout
   1. Data augmentation many forms cropping etc
   2. Drop out sets neurons to zero - effect of approximately averaging over an exponential number of smaller sub-networks, helps to work against co-adaptation of feature detectors
      1. Downside only really good for fully connected layers, not convolutional layers
6. **Cutout viewed close to drop out, but with spatial stuff applied prior - so visual features removed from the subsequent feature maps. - more accurately though cutout can be viewed as data augmentation**
7. Implementation
   1. They did targeted first, found random had about same performance for much less work

**Mixup**

1. Similar motivations - but this time very sensitive to adversarial examples
2. Mixup can almost be thought of as naive linear adversarial training - also helps with over fitting because encourages the idea of linearity in the models network
3. Can be thought of as a data augmentation method to encourage your model to behave linearly in between training examples - leads to linear decision boundaries **SO PGD attack is feeding it linear examples that it is already trained on - and nice because doesn’t add in significant computational overhead**

**Auxiliary rotation head**

1. Their work showed self supervision can introduce robustness in various ways
2. The rotation head can increase robustness - leads to better beformance against PGD and corrupt data - we struggled with this implementation though, so for some reason did not do super hot