- Weak supervision : Using noisy or imperfect labels for training on large data to reduce cost and increase efficiency of human effort.
- Inpainting : process of reconstructing lost or deteriorated parts of images and videos.
- Auto Augment : Choose best policy based on max validation accuracy.
- Different augmentations are used for train & test : while training, different random size RoCs are taken to increase training data, but while testing, centre crop is used which is biased towards centre important images. So train and test distributions are from different RoCs.
- Fine-tuning two layers is sufficient to compensate for distribution shift caused due to different size crops.
- Analysis shows that increasing crop size while testing compensates for randomly cropping RoCs while training. Training at low resolution and testing at higher resolution also improves efficiency.
- Training at low resolution makes more GPU available for larger batches which improve optimization.
- During testing, either TTA is applied or multi-scale input is used to make it scale-invariant.
- Frequency distributions of % area of image cropped by standard data augmentations show that test and train resolutions are very different.
- P-pooling was used to adapt the network to higher test resolution.It is the exponential average of the pixels to power p. GeM is a generalization of the average pooling commonly used in classification networks ($p = 1$) and of spatial max-pooling layer ($p = \infty$).
- P-pooling is a balanced pooling intermediate of max and average pooling used in semantic segmentation. Max pooling adds non-linearity and invariance which decreases sensitivity of model for segmentation, while average pooling is linear but retains details of image.
- Significant effects of RoC sampling :
    - Change in apparent size of objects in image due to resizing. CNNs are not robust to scale changes of objects.
    - Different crop sizes change statistics of activations of the CNN mainly after global pooling layers.
- CNN is hypothesised to learn scale invariance by training on randomly scaled objects. But in standard pre-processing, size distributions vary for train and test.
- Relation between apparent and actual object sizes : apparent size is inversely proportional to distance between object and camera.
- While training if RandomResizeCrop is used, then apparent size of object is standardized (independent of input image size). It is important as CNN don’t have inbuilt scale invariance. Also centre crop and resize at test time standardizes apparent size.
- But the ratio of apparent sizes of train and test objects is dependent on a random variable. The expectance of ratio is approximately 0.80.

- Convolutional layers statistics don't change by changing input size as the receptive field is same, but receptive field changes for global average pooling layers hence changing activation statistics.
- Activation statistics of average pooling show that if test resolution is smaller, then more chances of zeros and distribution is sparse and spread out. At higher resolution, distribution is less sparse and less spread out.
    - This observation shows that with change in test resolution, there is chance of getting output outside range of training output values.
- Even then increasing test resolution improves accuracy as it probably preserves apparent size and prevents train-test size mismatch.
- As expectance of ratio of test to train apparent sizes is 0.80, increasing test resolution by 1.25 times balances it. It is better to increase resolution of image than just zooming the crop as we may miss out some parts. This may skew distribution of activation statistics as observed.
- Approaches to compensate for skewed distribution:
    - Parametric adaptation: New layer is added after average pooling layer to balance activation distribution by a scalar transformation.
    - Adaptation by fine-tuning: Fine-tune last layers on same training set with increased train resolution. Batch-norm layers before average pooling layers are also fine-tuned as the network needs to adapt to sparsity. TTA is also used to prevent more domain shifts.
- Results showed that best accuracy of all models were obtained at increased test resolution.
- The higher the difference in resolution between training and testing, the more important is batch-norm fine-tuning to adapt to the data augmentation.
- It was observed that increased resolution is more efficient.
- RandomResizeCrop gave better results compared to resize with fixed size random crop.
- Training the classifier along with Batchnorm layer with TTA gave best result than any other combination.
- Adapting distribution before last ReLU before spatial pooling is effective in adapting to sparsity.
- The output of global pooling ressembled extreme value distribution, so it could be fit by a frechet distribution. Frechet distribution with new K_{test} is transformed to old distribution by a scalar transformation, which is used as activation after pooling layer.
-