

Improving your neural network

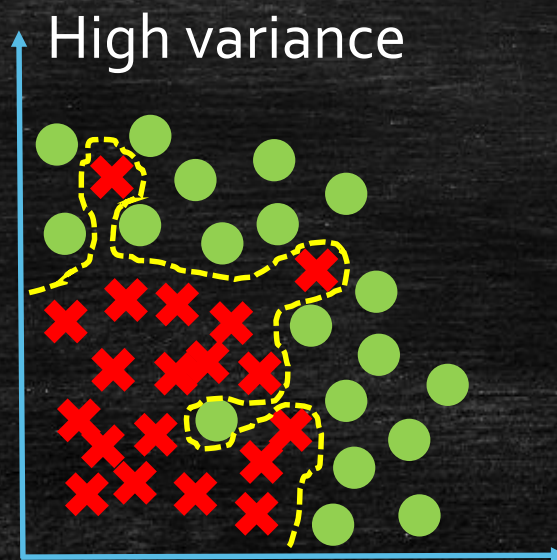
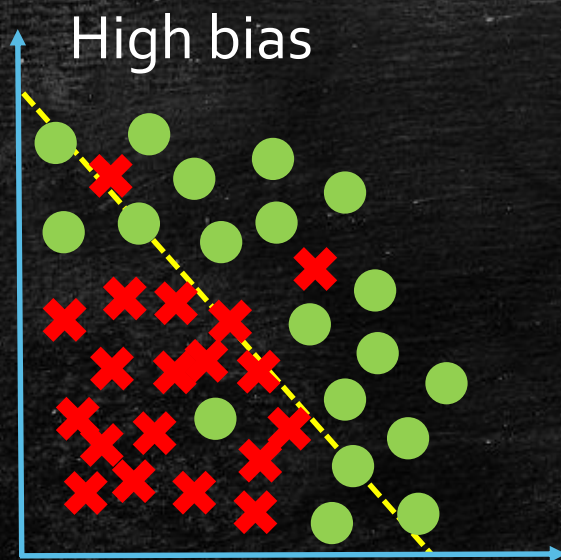
Joe Yeh, M.D.

Factors to consider when trying to improve neural network

- Data quality
- Data partitioning
- Hyperparameter tuning
- Regularization
- Data augmentation

What is a good neural network model?

- A model that has low bias and low variance.



Parameters vs Hyperparameters

- A machine learning **model** is the definition of a mathematical formula with a number of parameters that need to be **learned** from the data. That is the crux of machine learning: fitting a model to the data.
- Some parameters express “higher-level” properties of the model such as its complexity or how fast it should learn. They are called **hyperparameters**. Hyperparameters are usually fixed before the actual training process begins.

Hyperparameters

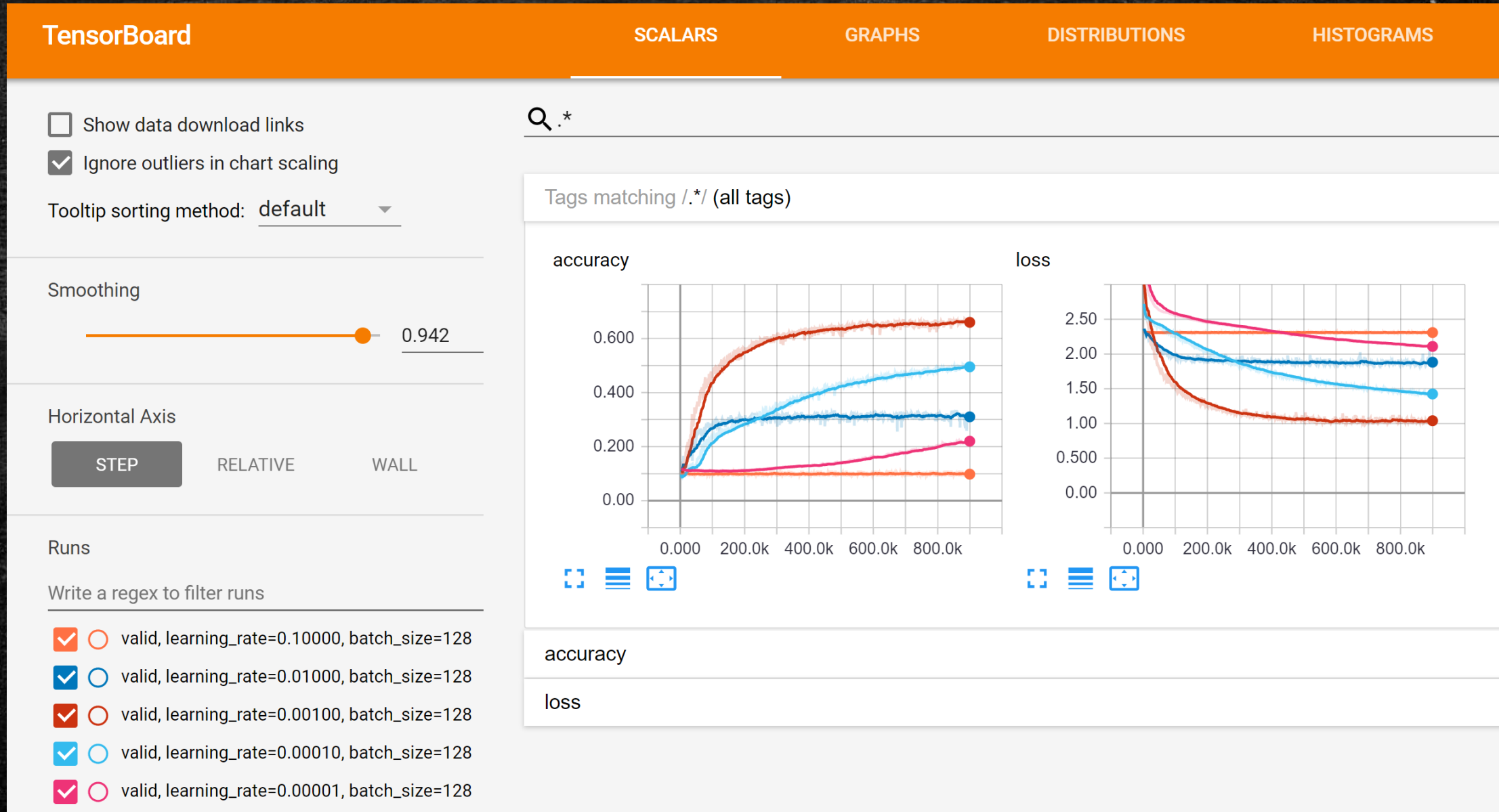
- Learning rate
- Learning rate decay rate
- Parameters of optimizers
- Number of hidden layers
- Number of hidden units
- Mini-batch size

Important rules in hyperparameter tuning

- It's very difficult to get an intuition on appropriate hyperparameters in advance, so the key is to explore hyperparameters space efficiently.
- Begin with a small subset of data if you have a very large dataset.
- Change hyperparameters one at a time.

Choosing appropriate learning rate

Example:
CIFAR-10
7-layer CNN
Adam Optimizer



Choosing Minibatch Size

Example:
CIFAR-10
7-layer CNN
Adam Optimizer

TensorBoard

SCALARS

GRAPHS

INACTIVE

☐ Show data download links

☒ Ignore outliers in chart scaling

Tooltip sorting method: default

Smoothing

0.9

Horizontal Axis

STEP

RELATIVE

WALL

Runs

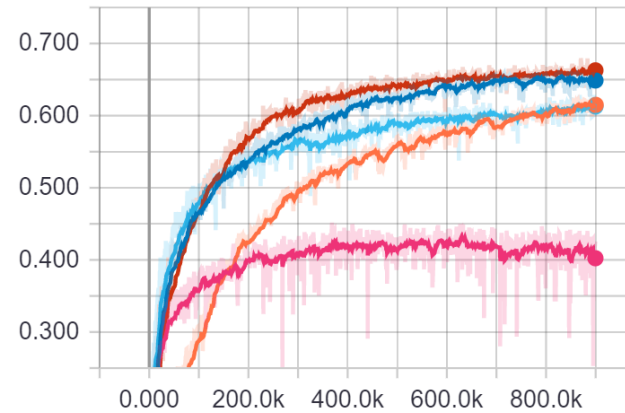
Write a regex to filter runs

- ☒ ☐ valid, learning_rate=0.0010, batch_size=512
- ☒ ☐ valid, learning_rate=0.0010, batch_size=128
- ☒ ☐ valid, learning_rate=0.0010, batch_size=32
- ☒ ☐ valid, learning_rate=0.0010, batch_size=8
- ☒ ☐ valid, learning_rate=0.0010, batch_size=2

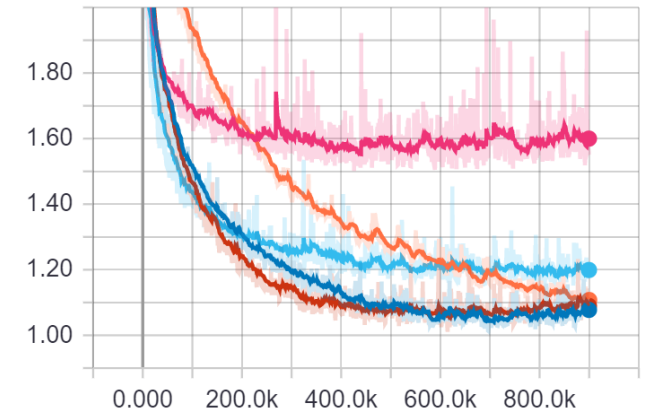
Q.*

Tags matching /.*/ (all tags)

accuracy



loss



accuracy

loss

Methods for Hyperparameter Tuning

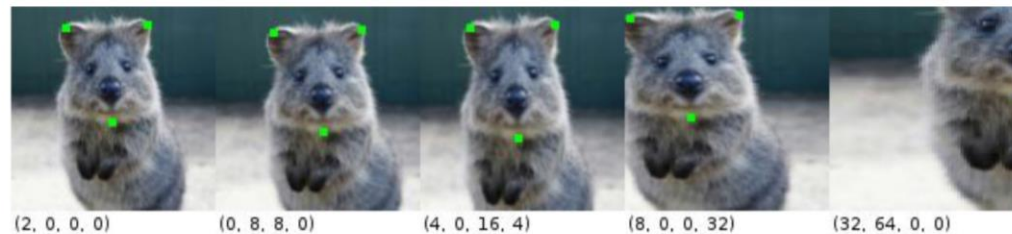
- Grid search
 - Curse of dimensionality
- Random search
 - Lack of *a priori* knowledge
- Bayesian optimization
 - <https://cloud.google.com/blog/big-data/2017/08/hyperparameter-tuning-in-cloud-machine-learning-engine-using-bayesian-optimization>
 - <http://papers.nips.cc/paper/4522-practical-bayesian-optimization>

Data Augmentation

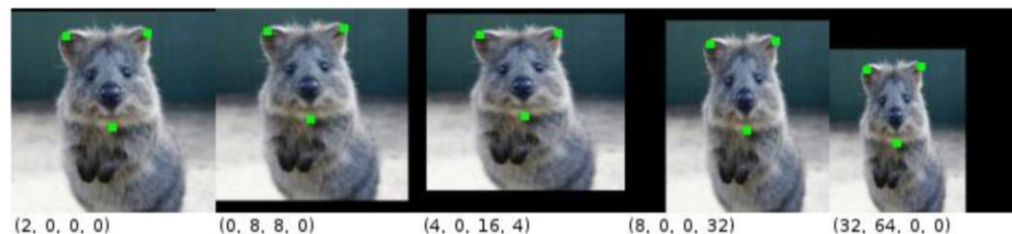
- Apply some form of transformation to the training data (but not label) to increase the size of effective training set.
- Regarding CNN for image recognition, it means:
 - Using image processing to manipulate input image to increase effective training set

Some examples of data augmentation

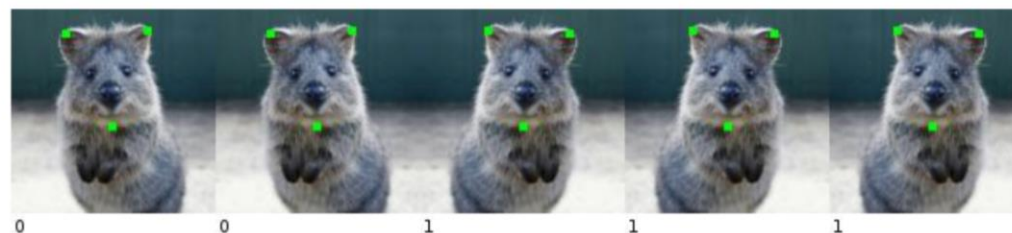
Crop
(top, right,
bottom, left)



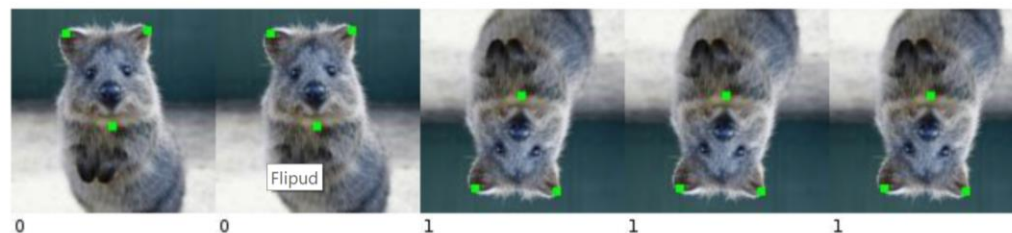
Pad
(top, right,
bottom, left)



Fliplr

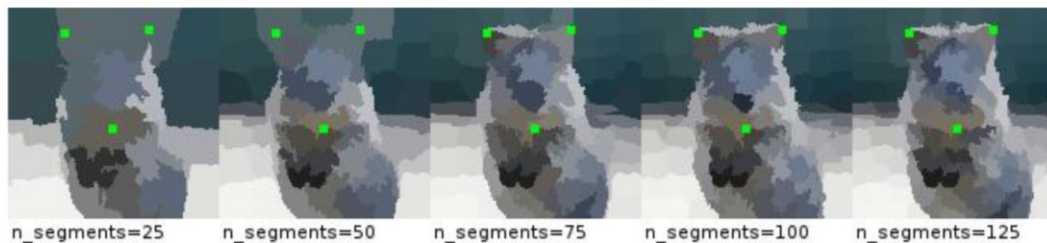


Flipud

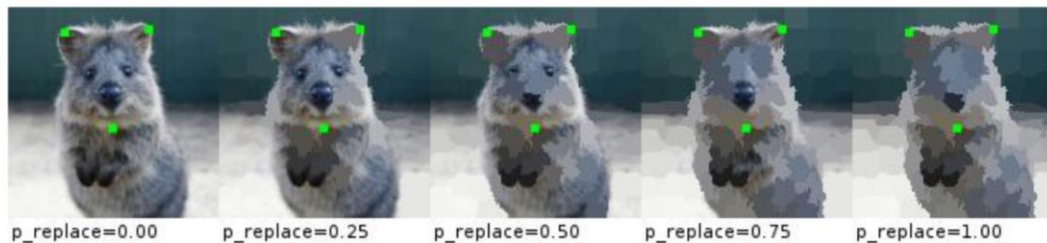


Some examples of data augmentation

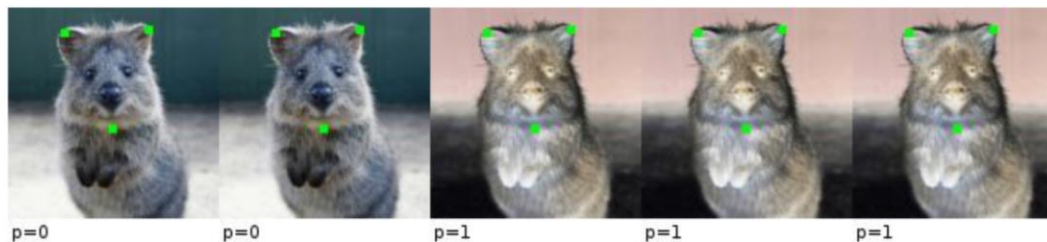
Superpixels
p_replace=1



Superpixels
n_segments=100



Invert



Invert
(per_channel)

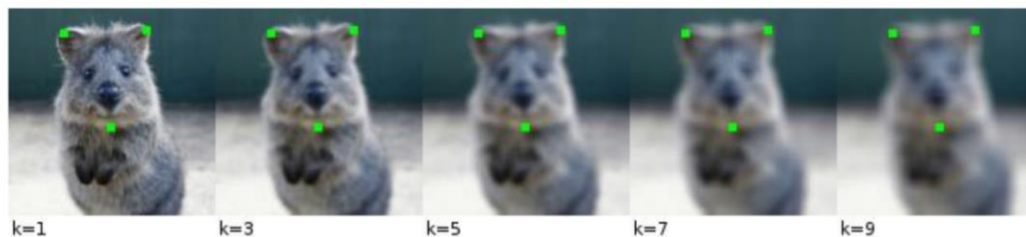


Some examples of data augmentation

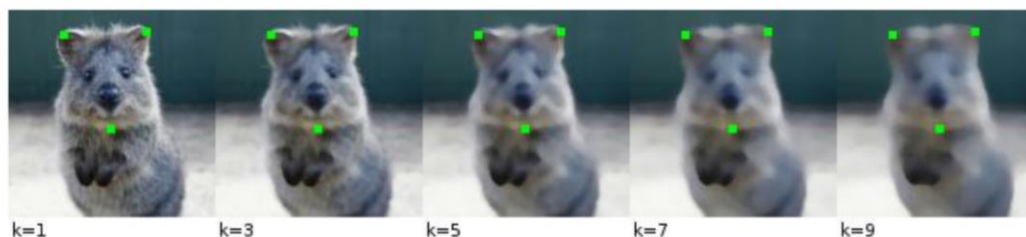
GaussianBlur



AverageBlur



MedianBlur

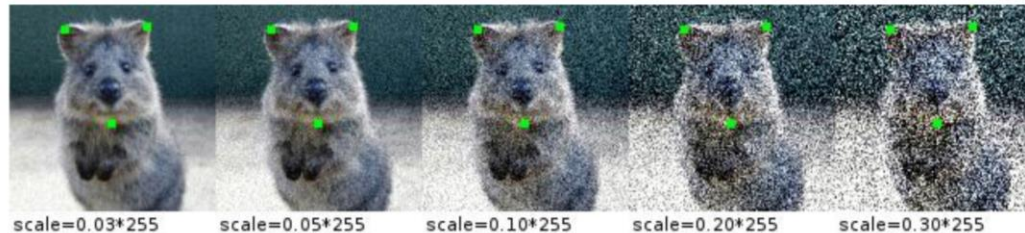


BilateralBlur
sigma_color=250,
sigma_space=250

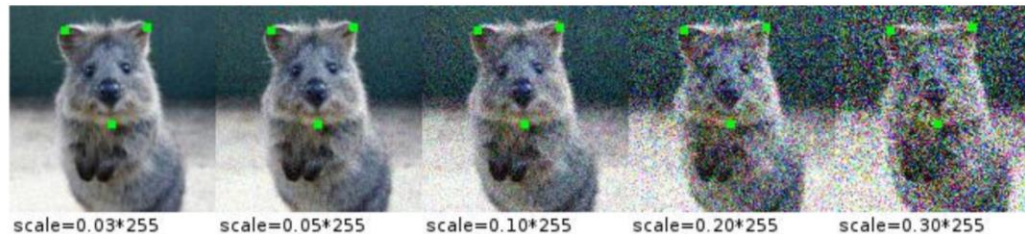


Some examples of data augmentation

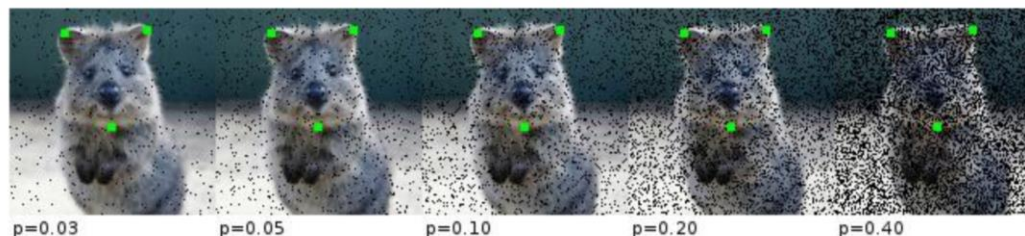
AdditiveGaussianNoise



AdditiveGaussianNoise
(per channel)



Dropout



Dropout
(per channel)

