Data Augmentation

Andrew H. Fagg

Data Augmentation

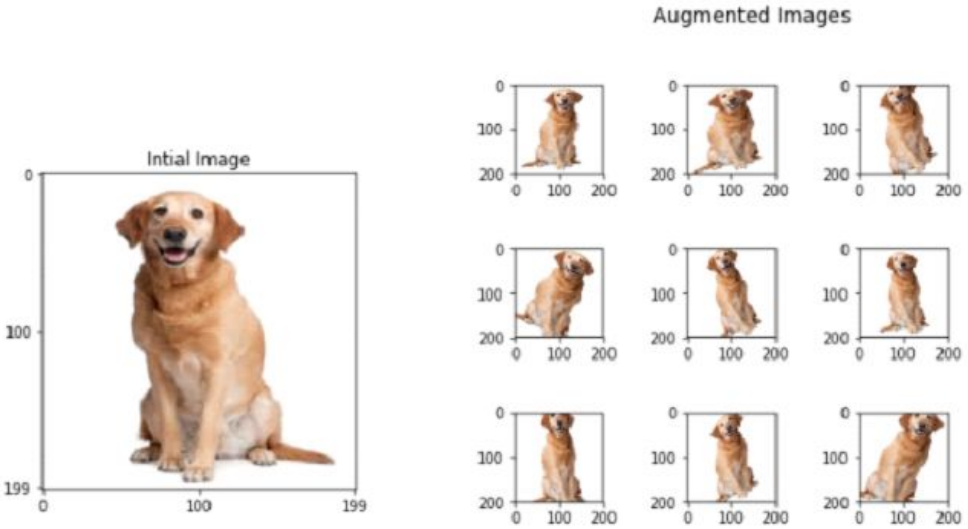
- Often training data sets are too small to provide an adequate sampling of the universe of data samples
- Would like to augment the data set in some rational way
- One possibility:
  - Take a real sample from the training data set and alter its input feature vector in some small way
  - This change in feature vector is small relative to the full distribution of the data, so we feel comfortable about copying the class label from the original point

Image Augmentation

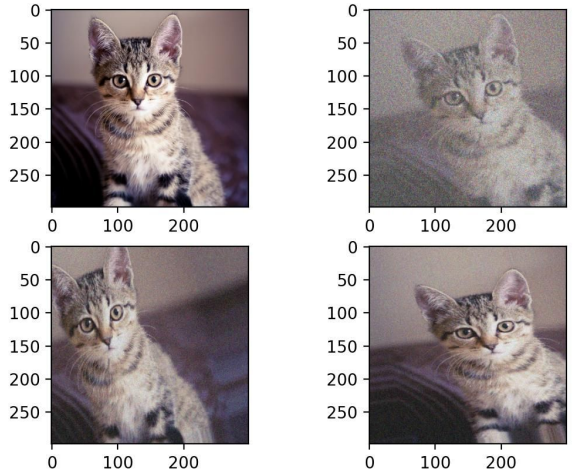
What does this look like for images? If we start with an image of a dog, do any of these change our class label?

- Flip the image along a vertical axis
- Flip the image along a horizontal axi
- Shift the image left/right or up/down a little bit
- Rotate the image a little bit
- Add random noise to the image

Image Augmentation



medium.com



Stackoverflow.com: Andriy Makukha

Image Augmentation

What does this look like for images? If we start with an image of a dog, do any of these change our class label?

- Flip the image along a vertical axis
- Flip the image along a horizontal axi
- Shift the image left/right or up/down a little bit
- Rotate the image a little bit
- Add random noise to the image

These changes are mostly small enough to not change the label

ImageDataGenerator

The ImageDataGenerator class does a lot of this work for us

- horizontal_flip
- vertical_flip
- height_shift_range
- width_shift_range
- rotation_range
- shear_range
- brightness_range
- zoom_range
- fill_mode: {"constant", "nearest", "reflect" or "wrap"}

Using the ImageDataGenerator

- Start simple: only introduce a couple transformations at once
- Small changes to the images may be enough
- Only augment the training set
  - Never augment the validation or testing sets: we want these to not be corrupted (any more than they already are)