# SGN-26006 Advanced Signal Processing Laboratory

*Image Recognition Assignment*

## Background

In this assignment we familiarize ourselves with modern machine learning; in particular *deep learning*. Image categorization is probably the most studied application example of a deep learning. As an example of designing a deep neural network, let us consider the Oxford Cats and Dogs dataset [1], where the task is to categorize images of *cats* and *dogs* into two classes. In the original pre-deep-learning era paper, the authors reached accuracy of 95.4 % for this binary classification task. In this task, your goal is to exceed this accuracy with modern tools.

We use a subset of 3687 images of the full dataset (1189 cats; 2498 dogs) for which the head ground truth location is available. We crop a square shaped bounding box around the head and train the network to categorize based on this input. The bounding box is resized to fixed size $64 \times 64$ with three color channels. We choose the input size as a power of two, since it allows us to downsample the image up to 6 times using the maxpooling operator with stride 2.

We consider two approaches to network design:

1. Design a network from scratch (with the structure of Figure 1)
2. Fine tune the higher layers of a pretrained network for this task (use the VGG16 network as a basis).

Since the amount of training data is relatively small, the first option necessarily limits the network size in order to avoid overlearning. In the second case, the network size can be larger as it has been trained with a larger number of images before.

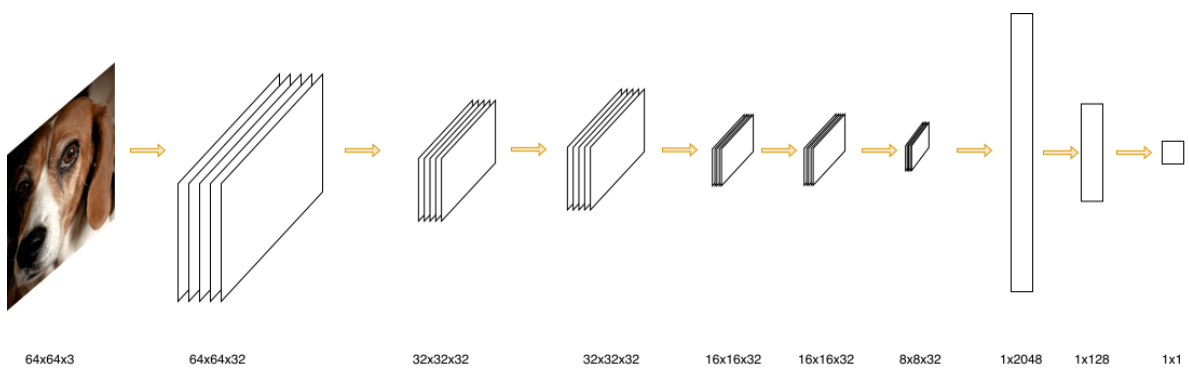


**Figure 1.** Topology of the small network.

## Tasks

Solve the following tasks, prepare a written report describing each item, and return in a zip file along with the code you implemented. It is recommended to use Keras + Tensorflow (or Theano), but also other deep learning frameworks are allowed (e.g., pyTorch). The use of MatConvNet is acceptable as well, but discouraged, as the importance of Python as the language of data science is increasing rapidly.

1. Download the cats and dogs dataset from http://www.robots.ox.ac.uk/~vgg/data/pets/. Extract the data and parse the cat/dog labels. Note that the annotations contain also fine grained classes (dog/cat breed), but we will focus on the two-class problem only.
2. Split the dataset into fixed training and testing sets (80% / 20%).
3. Design a convolutional network with the structure of Figure 1. Train the network for 50 epochs, store the classification accuracies and add a plot of accuracies to your report.
4. Design a fine-tuned network based on a large pretrained one.
    a. Instantiate the VGG16 network in Keras (or the platform of your choice). In keras, this is available at `keras.applications.vgg16`. The platform automatically downloads the network weights trained with the 1.2 million sample imagenet dataset.
    b. This network is trained with 1000 classes while our data has only 2. To correct this, remove the last 1000-node layer from the existing net, and substitute it with a 2-node layer initialized at random. Train the network for 50 epochs, store the classification accuracies and add a plot of accuracies to your report.

## Instructor

The instructor is Associate Prof. Heikki Huttunen (Heikki.Huttunen@tut.fi). Return the report and code to Moodle by the date given at the course Moodle page.

## References

[1] Parkhi, O. M., Vedaldi, A., Zisserman, A. and Jawahar, C.V., "Cats and Dogs", *IEEE Conference on Computer Vision and Pattern Recognition*, 2012.