Asdf

* Sparse and unevenly distributed data base:
* 9,500 labeled images of 4,500 whales as training data  
  - 2,000 whales, where only one image is available  
  - 1,000 whales with 2 images
* 15.000 images as test data
* Images differ in shape, resolution, color / grayscale

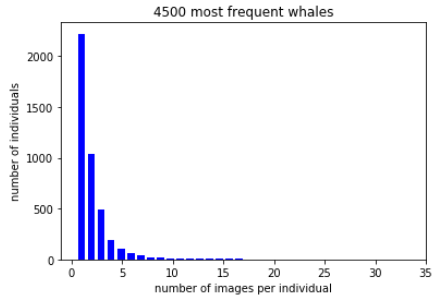
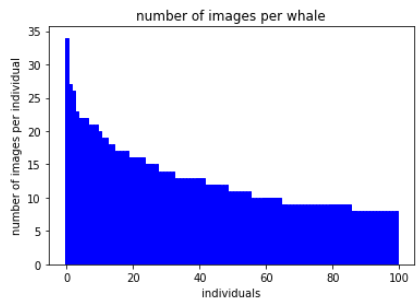
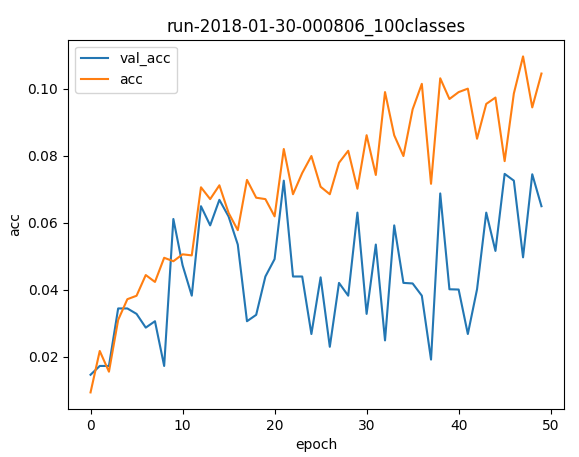
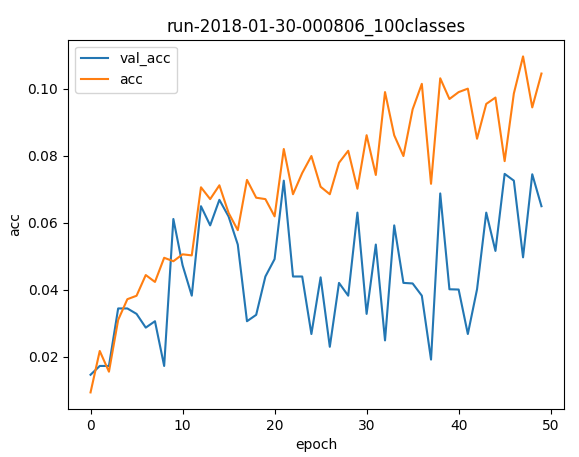
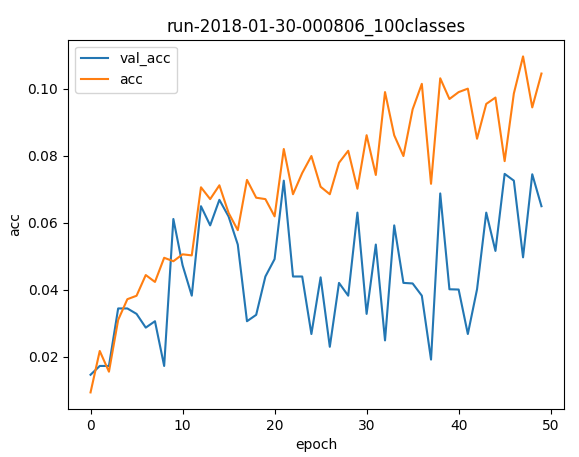
**Statistics**

**Humpback Whale Identification Challenge: a Kaggle competition**

Fabian Glöckle, Jochen Sautter, Benjamin Wilhelm

**The Challenge**

* 40 years of scientific whale monitoring generated a large collection  
  of images of tail flukes of Humpback whales
* 25,000 images were assigned to >4,000 individual whales by researchers
* Try to recognise the whales by their flukes with a deep learning model



* 4 plots with   
  - 3 or 4 models (bars for Accuracy / MAP)  
  - MAP / Accuracy bars for 3 / 10 / 50 / 200 / 1000 / 4250 whales   
  - plots with some learning curves  
  - plot with Hyperband results

**Results**

* Use large CNNs pretrained on image-net Database as provided by Keras library
* Add one or two dense layers for our classification task
* Freeze parameters of trained cnn base-models, and train on our task with dense layers
* Base Models used: InceptionV3, Xception, ResNet50
* After training of dense layers unfreeze the 2 top layer-blocks of the cnn base-model and train more epochs
* Use Data Augmentation: shifting, stretching, rotating
* Optimise Hyperparameter configuratoin using “HyperbandSter” framework (successive Halving combined with Bayesian optimisation)

**Fine-tuning**

**Our approach**

Asdf

