Individual project

- Coding or literature review?
- Own data or other data?
  - Replications or improvements to existing papers are a great option!
- Group projects are possible
- Scope: 4 weeks
- Products: 30 min presentation (exam week), submit code or essay/review

Past individual projects

- Segmenting trees in drone imagery
- Mapping Pika habitat in aerial imagery
- Identifying fireflies by their flash pattern
- Predicting pH preference of microbial species
- Species distribution models of beetles
- Predict carbon cycling from nutrient input
- Did this DNA come from a plasmid?

Today

- train-validate-test
  - 3 way split, validate split is for tuning
- It's all about the data
  - generalization
  - out-of-distribution vs out-of-sample
  - scope of inference (scale, extent)
  - test set leakage
- transfer learning & pre-trained models
  - code

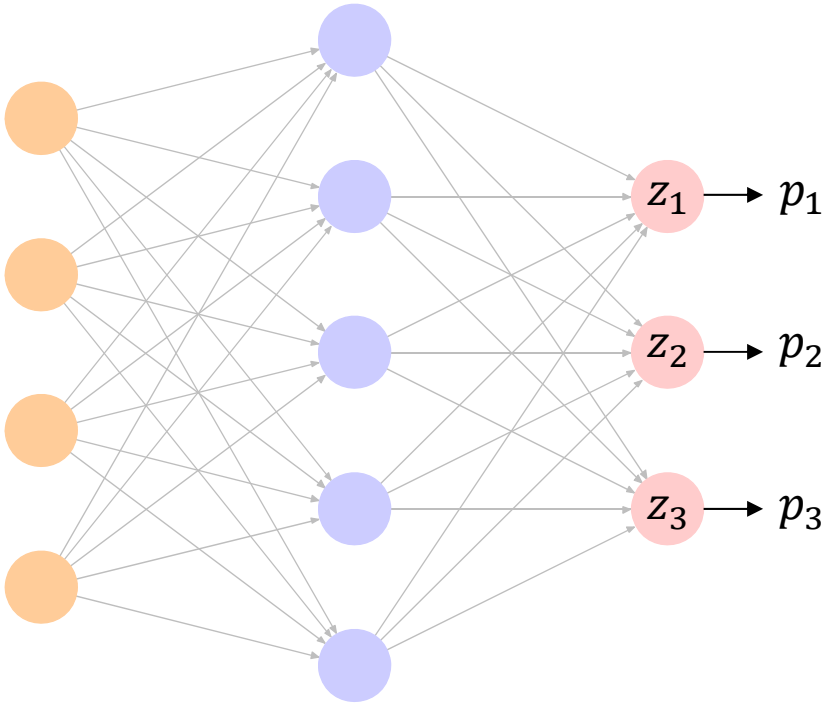
Softmax

Classification with multiple classes

Input *k* numbers (z_i) → output *k* probabilities (p_i)

Probabilities sum to 1

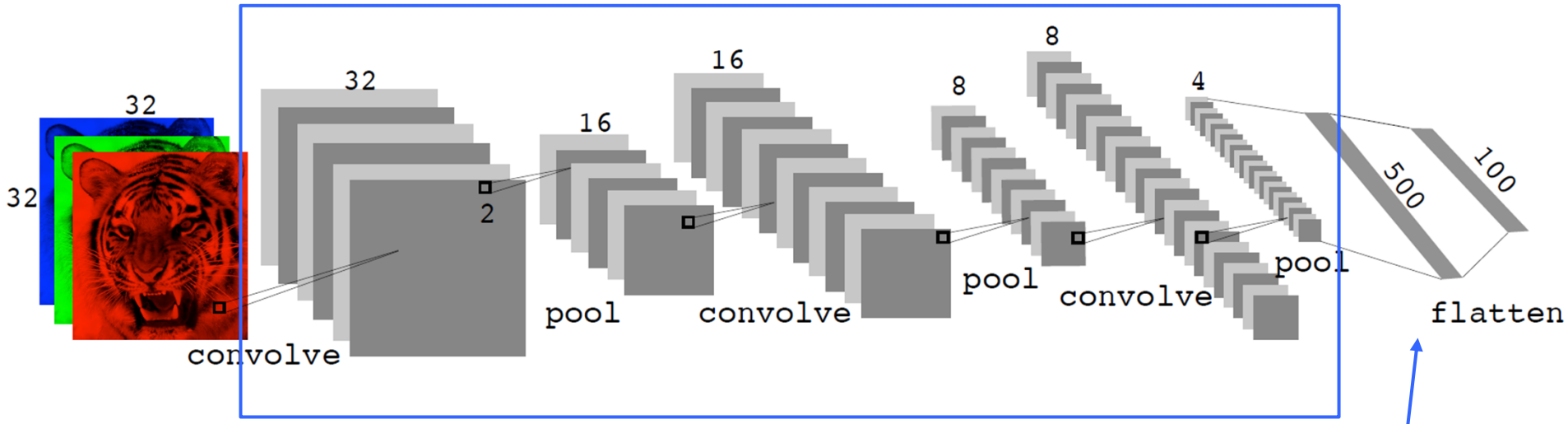
Exponential emphasizes the larger inputs



$$p_i = \frac{e^{z_i}}{\sum_{j=1}^{k} e^{z_j}}$$

Pretrained models

Transfer learning



Freeze convolve-pool layers (architecture, weights etc)

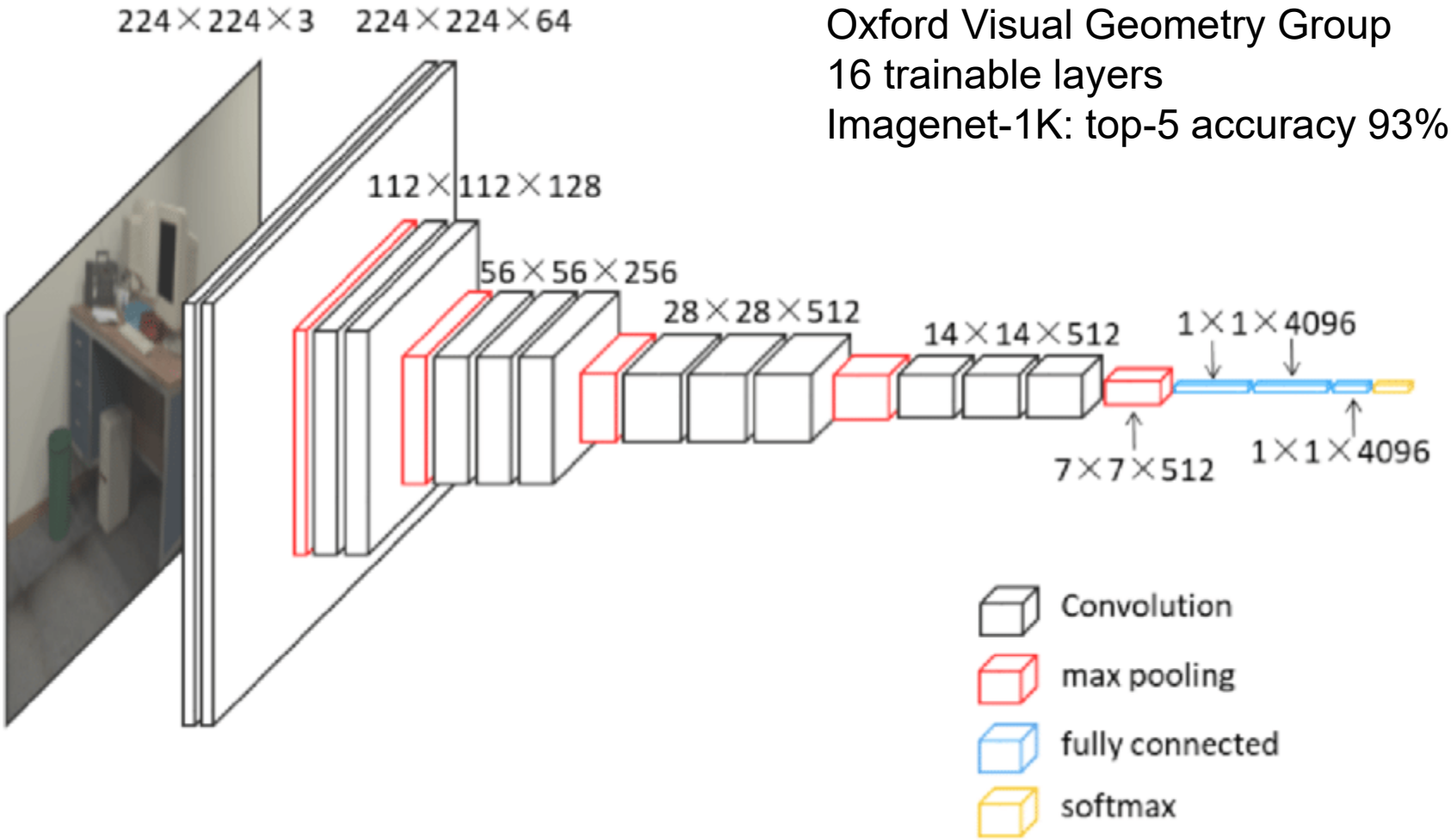
Reconfigure and/or retrain dense layers

Imagenet challenge

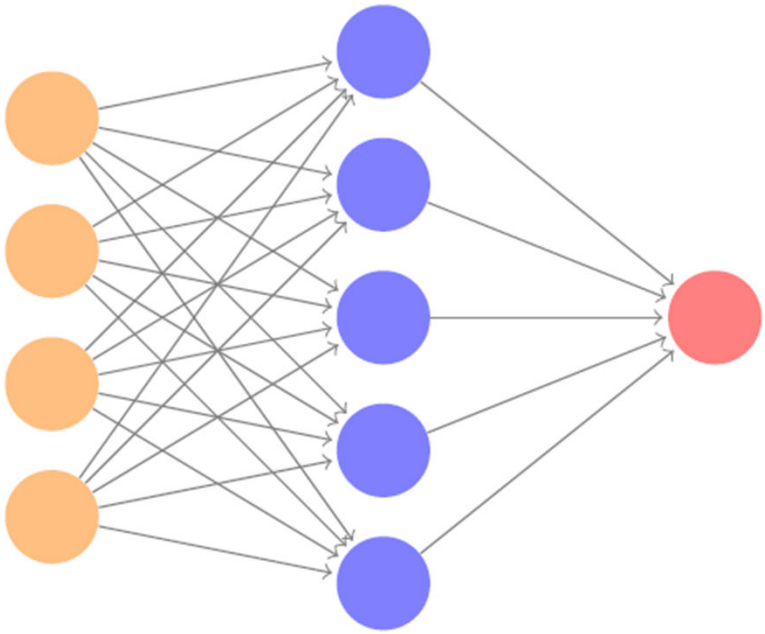
- Imagenet: Deng et al (2009)
  - massive collection of labelled images
- Imagenet-1K
  - 1000 classes, mean ca 1200 images each
  - multi-resolution
- Alexnet: Krizhevsky, Sutskever, Hinton (2012)
  - CNN, GPU trained
  - top 5 accuracy improved from 70% to 85%
- Current top-5 accuracy: 99%

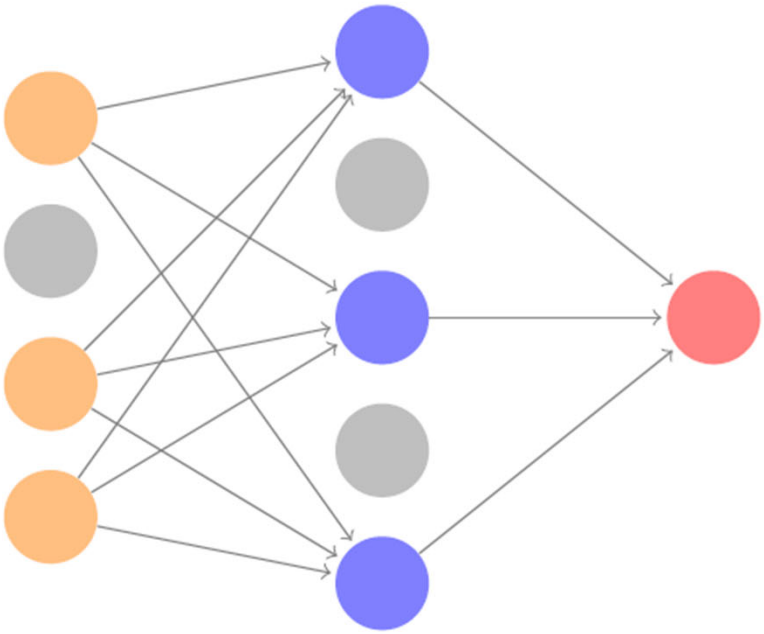
VGG16

Simonyan & Zisserman 2014
Oxford Visual Geometry Group
16 trainable layers
Imagenet-1K: top-5 accuracy 93%



Training algorithm: dropout





Like random forest sample columns