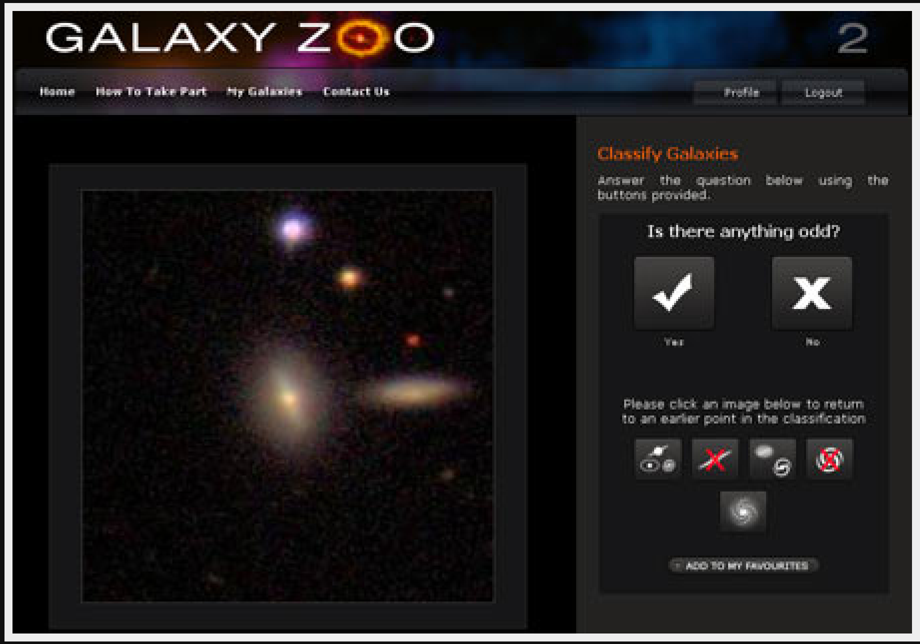
Galaxy Zoo Competition

Matthew Emery

https://lstmemery.github.io/kaggle-galaxy-presentation/#/

What Is Galaxy Zoo?

- All sky galaxy surveys produce more images than can be handled just by experts
- It's 2011, better crowd-source it



Galaxy Classification

- Over 900000 galaxies classified in a few months
- 37 categories arrived at by asking 11 questions
- After 40-50 users see a galaxy, the answers are aggregated and weighted vote fractions in the decision tree

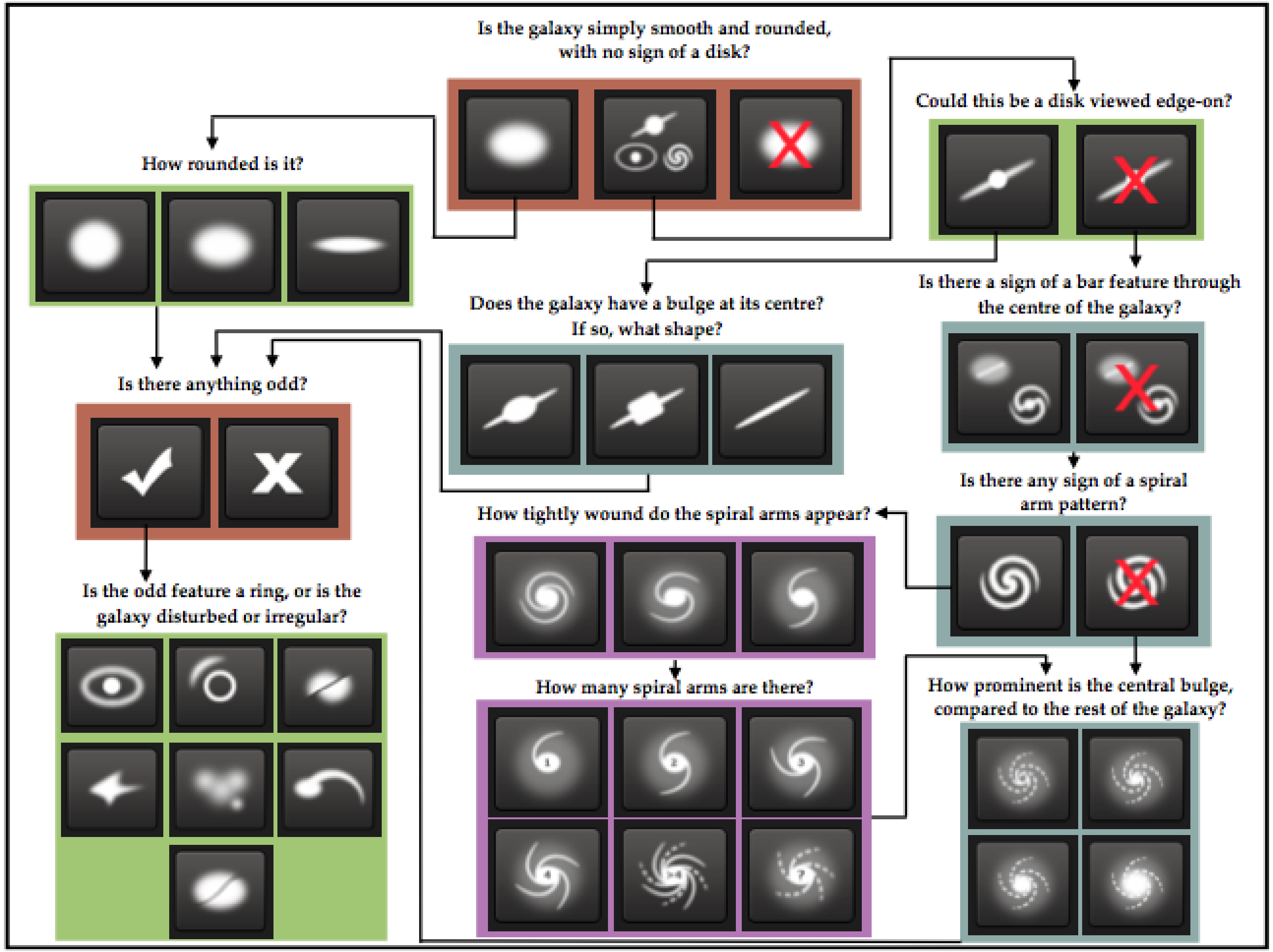


Figure 1. Flowchart of the classification tasks for GZ2, beginning at the top centre. Tasks are colour-coded by their relative depths in the decision tree. Tasks outlined in brown are asked of every galaxy. Tasks outlined in green, blue, and purple are (respectively) one, two or three steps below branching points in the decision tree. Table 2 describes the responses that correspond to the icons in this diagram.

Competition Details

Specifications

- Ran from December 20th, 2013 to April 4th, 2014
- 61,578 training images with vote fractions
- 79,975 test images
- 424 x 424 pixels
- Prize of $16,000

Loss Function

- This is a regression problem!
- The answers to the first question should sum to 1
- Answers to the next questions to sum to their parent probability

$$e(\hat{p}_k, p_k) = \sqrt{\sum_{k=1}^{37} (\hat{p}_k - p_k)^2}.$$

- Note: We are measuring error against the what the crowd answered. A “good” model will have the same biases as the crowd

The Winning Solution

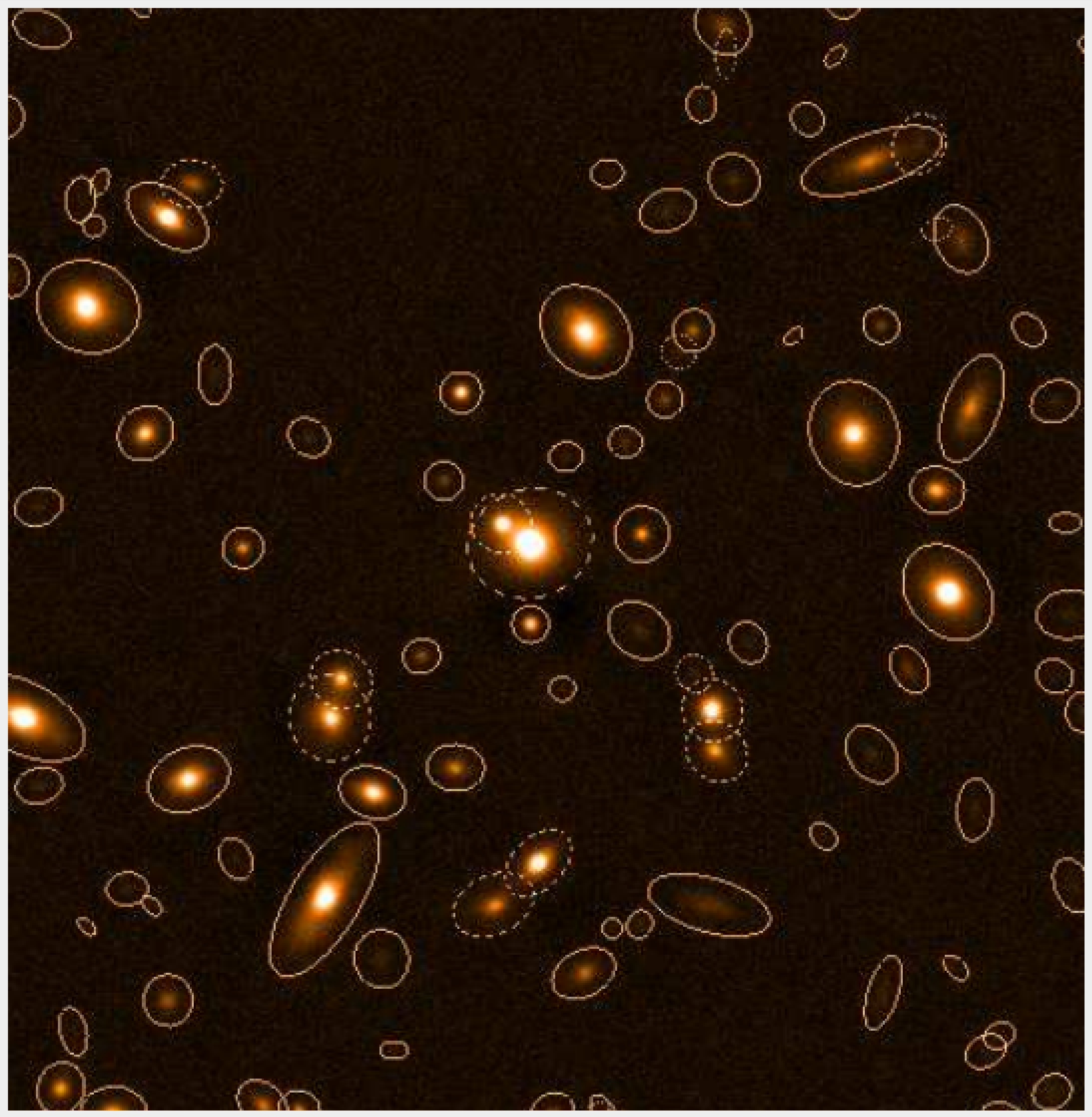
1st Place: Sander Dieleman

- Also won first place in the 2014 National Data Science Bowl with a team
- His approach was novel enough to get a paper in MNRAS and a job at DeepMind
- Co-author of Lasagne
- Second author on WaveNet, also on AlphaGo

Preprocessing

- Cropped 424x424 to 207x207
- Downscaled to 3x to 69x69
- Centering was done by Petrosian radius
- Normalized in some images but not others
- Keeping color significantly improved the model, despite it being artificial

- Avoided cropping out the object interest by centering with SExtractor (Source Extractor)



Data Augmentation

(a) none (b) rotation (c) translation (d) scaling (e) flipping (f) brightness

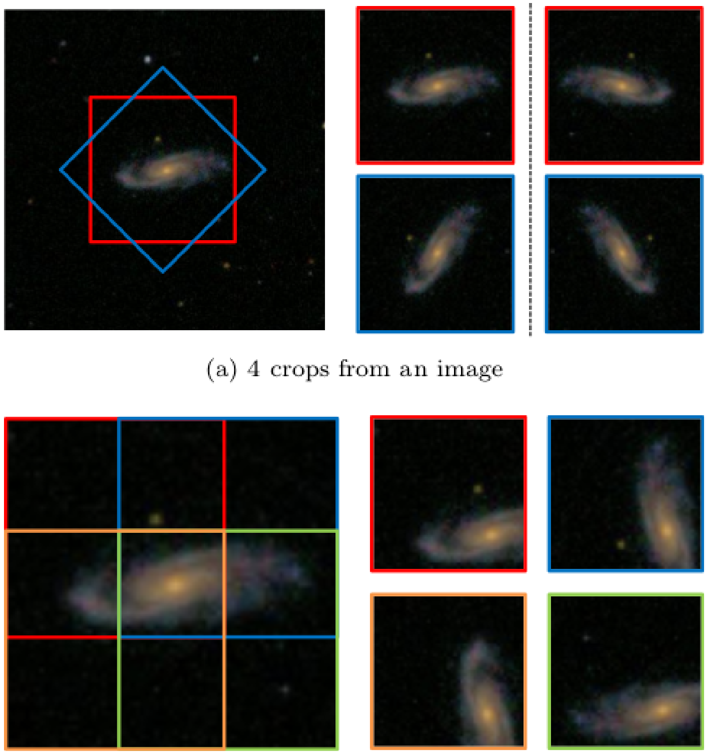
Figure 6. The five types of random data augmentation used in this model. Note that the effect of translation and brightness adjustment is fairly subtle.

- Rotating uniformly from 0 to 360
- Translating uniformly -4 pixels to 4 pixels
- Scaling log-uniformly from 1.3^-1 to 1.3
- Flip as a Bernoulli event with probability 0.5
- Color perturbation using an equation in the ImageNet paper
- This is all being done on the CPU while the GPU trains the network

Rotational Invariance

Step 1: Rotate image 45 degrees and flip both images (4 images)

Step 2: Crop each 67x67 image into 4 overlapping 45x45 images (4x4=16 images)



(a) 4 crops from an image

(b) 4 viewpoints from each crop

Figure 7. Obtaining 16 viewpoints from an input image. (a) First, two square-shaped crops are extracted from the image, one at 0° (red outline) and one at 45° (blue outline). Both are also flipped horizontally to obtain 4 crops in total. (b) Then, four overlapping corner patches are extracted from each crop, and they are rotated so that the galaxy centre is in the bottom right corner of each patch. These 16 rotated patches constitute the viewpoints. This figure is best viewed in colour.

Network Architectures

Best Model

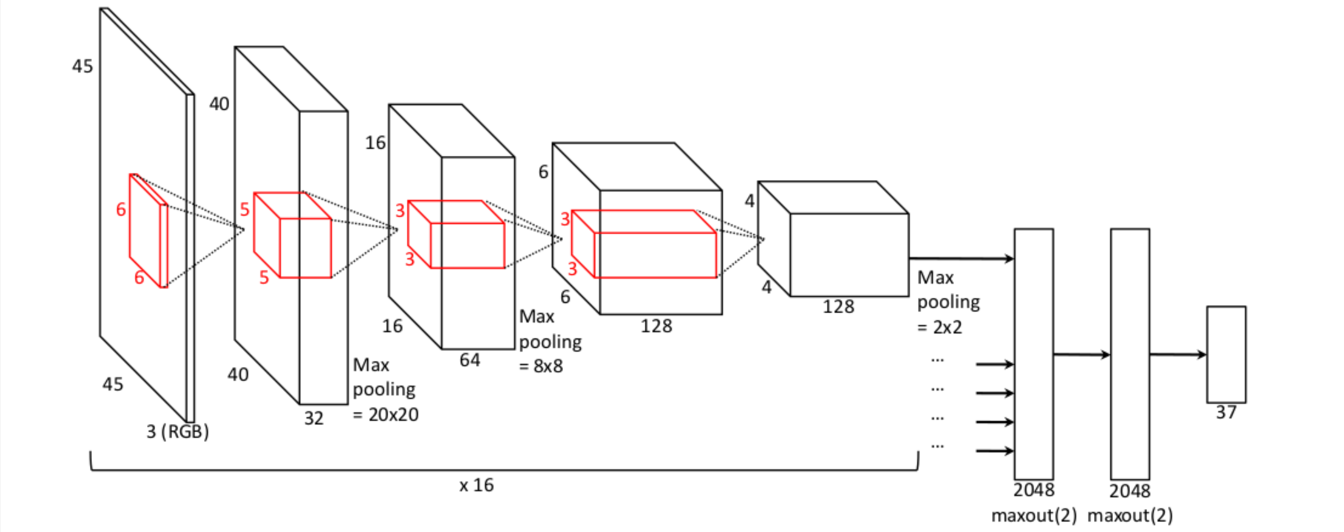


Figure 8. Schematic overview of the architecture of the best performing network that we trained. The sizes of the filters and feature maps are indicated for each layer.

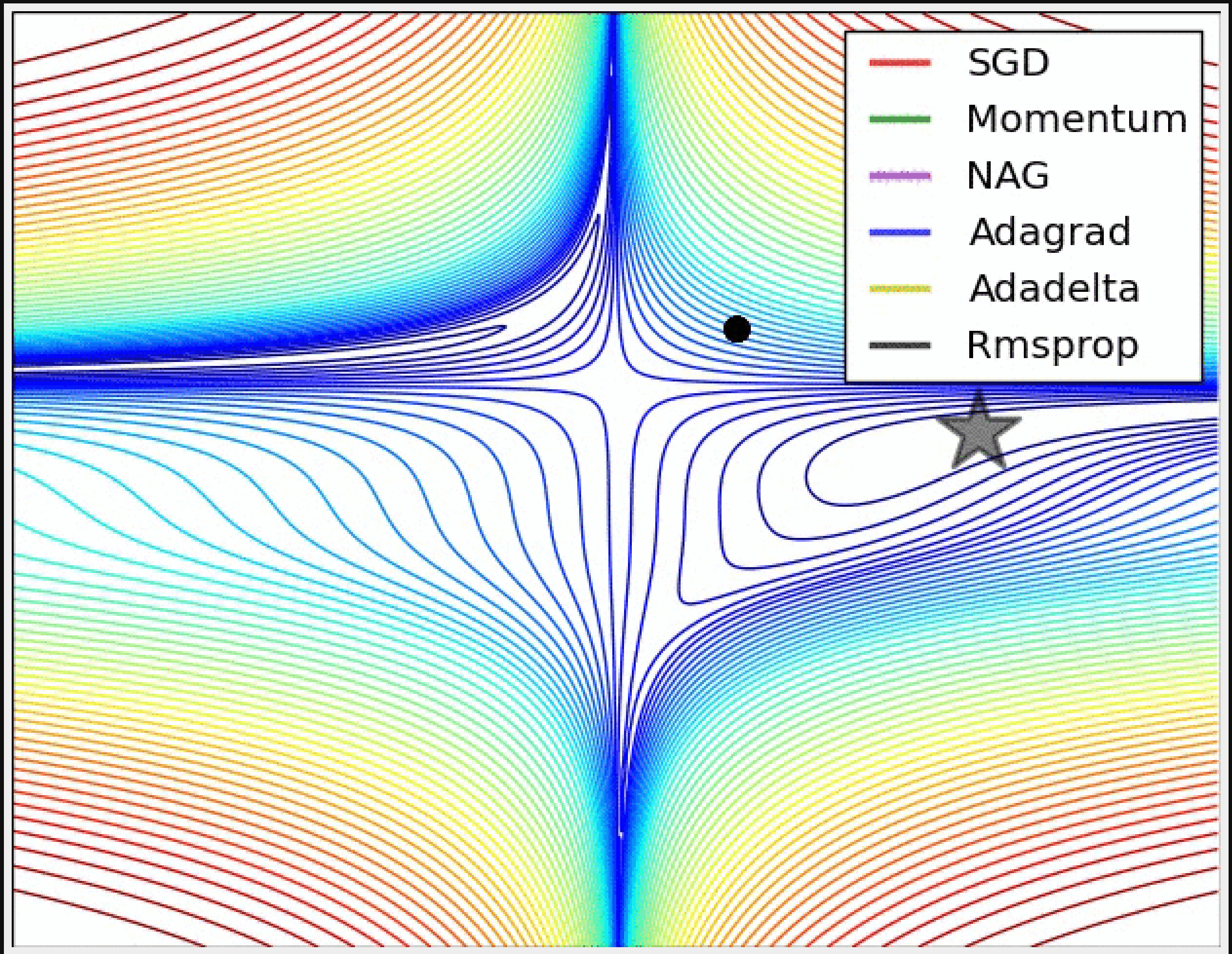
- All 16 viewpoints are fed in at the same time to maximize parameter sharing
- Best single model has 4 square convolutional layers (6-5-3-3)
- More than 100 models were tested, 17 were included in the final ensemble
- Maxout are like group-wise ReLUs
- The 37 were scaled to their constraints

Variants

- a network with two dense layers instead of three (just one maxout layer);
- a network with one of the dense layers reduced in size and applied individually to each part (resulting in 16-way parameter sharing for this layer as well);
- a network with a different filter size configuration: 8/4/3/3 instead of 6/5/3/3 (from bottom to top);
- a network with centered and rescaled input images;
- a network with a ReLU dense layer instead of maxout;
- a network with 192 filters instead of 128 for the topmost convolutional layer;
- a network with 256 filters instead of 128 for the topmost convolutional layer;
- a network with norm constraint regularisation applied to the two maxout layers;
- combinations of the above variations.

Training

- 67(?) hours of training for the best model on a GTX 680
- Nesterov Momentum was used (16 “minibatches,” effectively 256 because of architecture)
- Learning rate of 0.04 (updated to 0.004 after 18M samples, then 0.0004 after 23M)
- Dropout was used during training to prevent overfitting


SGD
Momentum
NAG
Adagrad
Adadelta
Rmsprop

Model Averaging

- Modeled across 17 architectures (they are all available on GitHub)
- For each model, averaged predictions across 60 different transforms
- 10 rotations x 3 rescalings x 2 reflections
- It takes 4 hours to get a prediction from a single model

How Did He Do?

model	leaderboard score	
	public	private
best performing network	0.07671	0.07693
+ averaging over 60 transformations	0.07579	0.07603
+ averaging over 17 networks	0.07467	0.07492

Table 3. Performance (in RMSE) of the best performing network, as well as the performance after averaging across 60 transformations of the input, and across 17 variants of the network. Please refer to Section 3 for details on how the scores were computed.

- His single best network outperformed everything else

Things that Didn't Work

- Adding Gaussian Noise to the image
- Changing gamma
- Downsampling less
- Adding shearing to preprocessing
- RMSprop or adadelta

References

1. I. Goodfellow, Y. Bengio, and A. Courville, Deep learning. 2016.
2. S. Dieleman, K. W. Willett, and J. Dambre, “Rotation-invariant convolutional neural networks for galaxy morphology prediction,” Monthly Notices of the Royal Astronomical Society, vol. 450, no. 2, pp. 1441–1459, Apr. 2015.
3. [1]“benanne/kaggle-galaxies,” GitHub. [Online]. Available: https://github.com/benanne/kaggle-galaxies. [Accessed: 16-Dec-2016].