The University Of Sheffield.

Cracking Captchas

Jennifer Parak

Supervisor: Dr. Richard Clayton

Introduction & Background

Completely Automated Turing test to tell Computers and Humans Apart (CAPTCHAs) are widely used on the Internet to keep it safe from bots and spammers. CAPTCHAs present a challenge-response test, based on a hard Artificial Intelligence problem, which is supposed to be difficult to solve for computers, but easy to solve for humans, such as recognising characters, audio or images. This project focused on cracking image-based captchas whose security is based on the hard AI problem of image labelling.

Project Aims

- To examine machine learning approaches that have been used to classify images
- To point out problems and difficulties of cracking an image recognition CAPTCHA
- To implement some of these approaches on a very simple CAPTCHA
- To identify the features of a 'strong' image recognition CAPTCHA

How to Crack a CAPTCHA 101

STEP 1: Pre-training with a Sparse Autoencoder

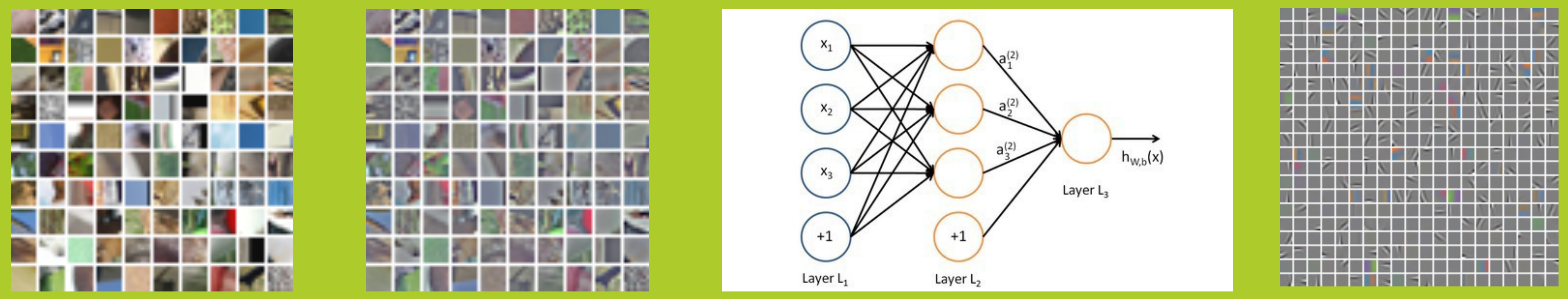
We train a Sparse Autoecoder, a simple Single-Layer Neural Network, to extract essential features from small patches sampled from unlabeled images [1]. The aim is to compute a representation of the input which can be reconstructed in the output layer. By putting a constraint on the hidden units, we force the unsupervised learning algorithm to compute a useful, compressed representation.

0.) Input: 8*8 Patches

1.) Pre-Processing

2.) Training using Backpropagation

3.) Output: Features



STEP 2: Fine-tuning with a Convolutonal Neural Network

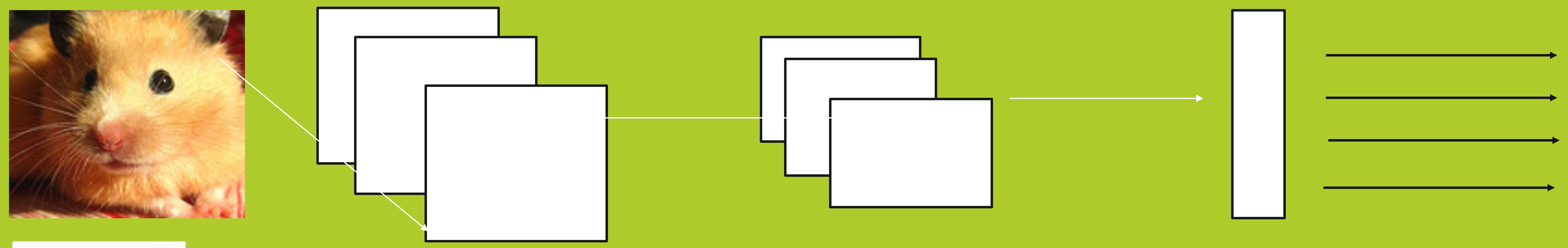
We developed a Covolutional Neural Network to learn features from our training set using our feature detector learnt by the Autoencoder. It consists of several layers: (1) The Convolutional Layer computes a representation by performing dot products between the local image regions and the filters learnt. (2) In the Pooling layer we perform mean-pooling to reduce dimensionality. Specifically we divide the feature maps into 4 regions and take the average activation of each. (3) The Fully-Connected Layer is our classification layer, which estimates the probability of input belonging to one of the existing classes.

0.) Input Layer: 64*64

1.) Convolution Layer : 64*64 * 400

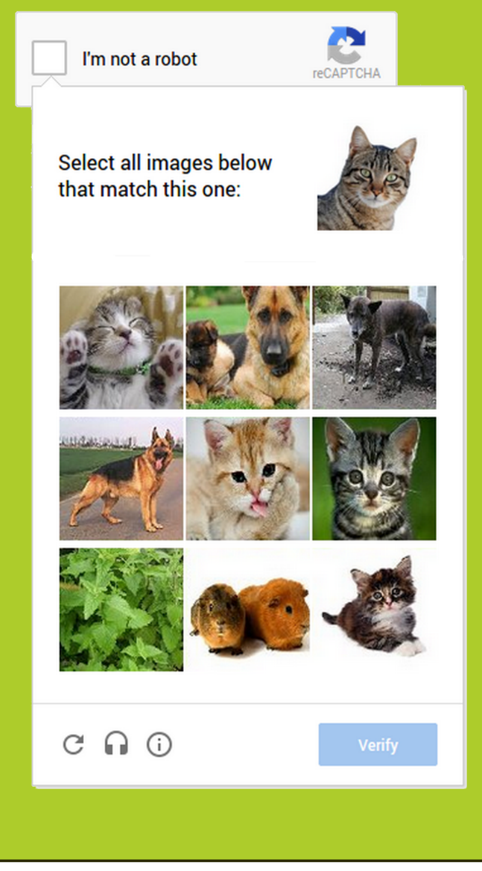
2.) Pooling Layer: 3 * 3 * 400

3.) Classification Layer: Predictions



STEP 3: Cracking CAPTCHAS

We tested our Convolutional Neural Network across the whole dataset to evaluate the accuracy, the main focus, however, was to explore whether our model could pass an image recognition CAPTCHA challenge. The testing process works as follows: The images that we want to predict are convolved and pooled and fed into the trained classifier. The classifier predicts the probabilities of the image belonging to one of the classes and outputs the class with the highest probability.



Select all images below that match this one:

0) Extract Images, 1) Convolution, 2) Pooling, 3) Classification

Results

Our test set included 3,200 images, 800 per class. The trained model achieves an accuracy of 70 % on the test set. Further experiments were conducted to explore the effect of pre-training on our results. We found that pre-training increases our results by 15%.

Conclusion

This project has successfully proven that we could use off-the-shelf approaches and technologies, in order to pass an image-based CAPTCHA challenge. Even with a relatively simple pre-trained Convolutional Neural Network, we managed to develop a system which is able to crack CAPTCHAs. Do image recognition CAPTCHAs still represent a hard AI problem? The development and training of a model capable of passing a CAPTCHA challenge is fairly involved and computationally expensive for networks that aim to achieve results comparable to human accuracy. Nevertheless, there are numerous libraries that facilitate the development of Convolutional Neural Networks.

To make an image recognition CAPTCHA more secure, we have identified the following three main features. Please note that implementing some of these features could also have a negative effect on the accessibility for human users.

1. Security based on Image Properties. We suggest CAPTCHA desginers to use large, distorted images, including multiple objects of specific categories.
2. Security based on Properties of the CAPTCHA challenge. To make the CAPTCHA challenges more secure, challenges have to be generated randomly and from a large, dynamic database consitsting of various different categories. To decrease the probability of a computer passing a CAPTCHA challenge, the number of rounds and images to be selected can be increased.
3. Security based on Limitations of Convolutional Neural Networks. There are objects which are easy and hard to classify by a Convolutional Neural Network. By using those images, the security of image recognition CAPTCHAs can be ensured. Furthermore, Neural Networks still struggle to interpret scenes or emotions the way that humans do. Recent research shows that Convolutional Neural Networks can be tricked into misclassifying images with a high confidence level, which are unrecognisable to humans [2].

References:

[1] Yaping Lu; Li Zhang; Bangjun Wang; Jiwen Yang, "Feature ensemble learning based on sparse autoencoders for image classification," in *Neural Networks (IJCNN), 2014 International Joint Conference on* , vol., no., pp.1739-1745, 6-11 July 2014

[2] Nguyen, A., Yosinski, J., & Clune, J. (2014). Deep neural networks are easily fooled: High confidence predictions for unrecognizable images. *arXiv preprint arXiv:1412.1897*.

Images used:

[3] **Google Bots.** Google reCAPTCHA [Online]. Retrieved from: https://www.google.com/recaptcha/intro/index.html. Last Accessed on 01/09/2015

[4] **IMAGENET Dataset.** Olga Russakovsky*, Jia Deng*, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, Alexander C. Berg and Li Fei-Fei. **ImageNet Large Scale Visual Recognition Challenge.** *IJCV,* 2015.

[5] **CAPTCHA Challenge.** Google reCAPTCHA [Online]. Retrieved from: https://www.google.com/recaptcha/intro/index.html. Last Accessed on 01/09/2015

Contact Details:
E-Mail: jenpaff0@gmail.com