

#12

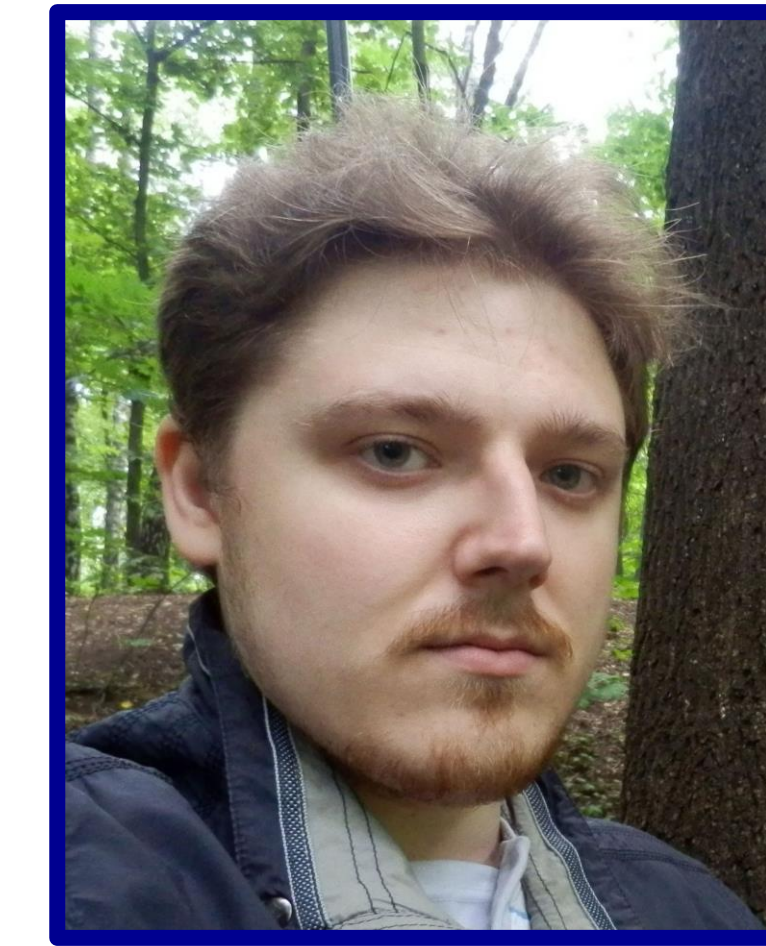


Preliminary Experiment on Emotion Detection in Illustrations Using Convolutional Neural Network.

Alexander N. Shtanko^{*1}, Sergey D. Kulik^{1,2}. *shtanko-mephi@yandex.ru

¹National Research Nuclear University MEPhI, Kashirskoe shosse 31, Moscow 115409, Russia,

²Moscow State University of Psychology and Education (MSUPE), Sretenka st. 29, Moscow 127051, Russia.



Alexander N. Shtanko



Sergey D. Kulik



SUMMARY

The paper describes an experiment on emotion detection in images. Usually, detection of emotions are performed on human faces so the algorithm can learn, for example, what a “happy human face” looks like. But can these algorithms transfer their understanding of happiness features onto different types of objects, like animals or cartoon illustrations. We, humans, can recognize and detect signs of emotions in new and unusual objects. Developing an algorithm capable of recognizing emotions in objects it wasn't trained on would allow for better human-like systems. This is a preliminary study on how well knowledge gained by a typical neural network detection system on a set of objects transfers to new, unknown objects. We collected small training datasets using cartoon illustrations of several animals of two categories: happy and sad. We tested the trained network on a set of illustrations depicting a different animal the network hasn't seen in training. The best performance achieved is 0.69 F1-score.

INTRODUCTION

The topic of this paper is object detection systems. One of the important targets of computer vision is emotions, because intelligent systems need to recognize and react to them. Humans can detect the signs of emotions in abstract shapes and inanimate objects (due to the human tendency to seek patterns), even in never seen objects (or in unusual representations of the objects). A system able to find areas expressing emotions without an attachment to specific objects it was trained on would be a step towards more human-like artificial intelligence.

APPROACH

In this paper, we're going to train a convolutional neural network on a dataset containing animal illustrations. The training goal is to locate emoting characters and classify expressed emotion into 2 categories: sad and happy. The performance of the network is then assessed on a testing dataset. However, emoting objects (cartoon animals) are different in training and testing dataset so we'll see how good the network understands emotion features rather than memorizing the emotion features of that specific cartoon animal. Training will be done on several datasets of different sizes and containing different animal types. Additionally, in the datasets there're illustrations of different styles which poses an extra challenge to the neural network.

Numbers of training examples in the datasets are relatively small because it's also important to test the learning ability on small datasets, considering in many applications it's hard to collect large amounts of training data.

METHODS

We used YOLO object detection system based on convolutional neural networks.

Out of various version of this system we used specifically YOLOv3-tiny because of its fast training and good results on small datasets.

To measure the performance of the networks we used a separate dataset, and counted the number of correct and incorrect detections given by the neural network and calculated precision, recall and F1-score. And we compare them between networks trained on different datasets.

RESULTS

We created five training datasets of increasing (from 1 to 5) variety of targets (target cartoon animal types) and dataset size. Targets (emoting characters) were marked by bounding boxes. Another dataset containing the target of different type was created to test the performance of the networks.

Next we trained neural network object detection system YOLO using YOLOv3-tiny architecture on each of the training datasets and tested the trained networks on the testing dataset.

Since the testing dataset contains different types of the target objects compared to the training datasets, it shows how well the neural network generalized the knowledge about the target emotions. The results are listed in Table 2.

Table 2. Networks performance on testing dataset.

Metric	Dataset 1	Dataset 2	Dataset 3	Dataset 4	Dataset 5
Class "sad": true positive	20	30	35	39	35
Class "sad": false positive	14	27	23	20	18
Class "happy": true positive	27	37	39	42	43
Class "happy": false positive	16	20	17	11	17
Precision	0.61	0.59	0.65	0.72	0.69
Recall	0.38	0.54	0.60	0.65	0.63
F1-score	0.47	0.56	0.62	0.69	0.66

ANALYSIS

In the first dataset there's only one type of object to learn from in the training dataset. The trained neural network performs pretty well when it comes to localization of the emoting targets, however classification is at a level of random guessing. Sometimes the system registers two detections for each class in the same position. With increasing the size of the dataset and including new types of animals, classification becomes better but still, the network makes a lot of mistakes.

On all these datasets network's recall is lower than precision which is typical for YOLO object detection.

Thus, the best performance we managed to achieve on new, unseen types of objects is 0.69 F1-score.

DISCUSSION

The results were somewhat successful given the variety in drawing styles and small size of datasets. However, multiple issues were encountered during research.

First of all, in this case we're still tied down to faces expressing emotions, in a sense it's no different from two stage (detection → classification) detection. If we want to approach generalized emotion detection, we should strive away from faces specifically to abstract emotions.

On the plus side, using illustrations allow to generalize emotion recognition beyond human face on real life photographs.

Another problem is that the emotions are subjective especially if the image is abstract.

Perhaps one of the ways to improve upon emotion detection is image segmentation instead of using bounding boxes.

CONCLUSIONS

1. We investigated the ability of object detection system YOLO to detect emotions using cartoon illustration images.
2. We collected necessary images for datasets, trained neural networks and measured their performance.
3. Depending on the training dataset the performance of the network went from random guessing to 0.69 F1-score.

ACKNOWLEDGMENTS

This work was supported by Competitiveness Growth Program of the Federal Autonomous Educational Institution of Higher Education National Research Nuclear University MEPhI (Moscow Engineering Physics Institute).