Automatic Weapon Detection in Social Media Image Data Using a Two-Pass Convolutional Neural Network

Jens Elsner
Thomas Fritz
Laura Henke
Oussama Jarrousse
Mathias Uhlenbrock
Stefan Taing
Munich Innovation Labs, Germany[1]

Abstract
Police analysts are faced with a deluge of data when monitoring the activities in specific areas of social networks and other internet data sources. Image recognition can help to prioritize the reading and subsequent analysis. The paper presents a case study for weapon detection in image data that has the potential to reduce the workload of the analyst by a factor of 200.

Keywords: Image Classification, Weapon Detection, TensorFlow, Social Network Analysis

1. Introduction

Police analysts are faced with a deluge of data when monitoring the activities in specific areas of social networks and other internet data sources. Image recognition can help to prioritize the reading and subsequent analysis. For example, when monitoring online resources of potential radicals, any posting of a weapon is of interest as it might indicate a possible threat. In recent years, image object classification using deep learning techniques has made significant progress with the advent of powerful computational architectures such as Graphical Processing Units (GPUs). The purpose of this paper is to study the performance of the application of the publicly available and open source TensorFlow framework (Abadi et al., 2015) to the problem of weapon recognition in images.

A classification approach that allows to incorporate and learn from analyst feedback using supervised learning while keeping the total retraining time of the classifier at a minimum is presented.

1 Corresponding author's email: je@munich-innovation.com

2. Methods

2.1. 2-Pass Image Object Detector

The presented 2-pass image object detector consists of two modules: First, the Search-Net, a region-based fully convolutional network (R-FCN) (Dai et al., 2016) with a ResNet-101 feature extractor (He et al. 2016) for

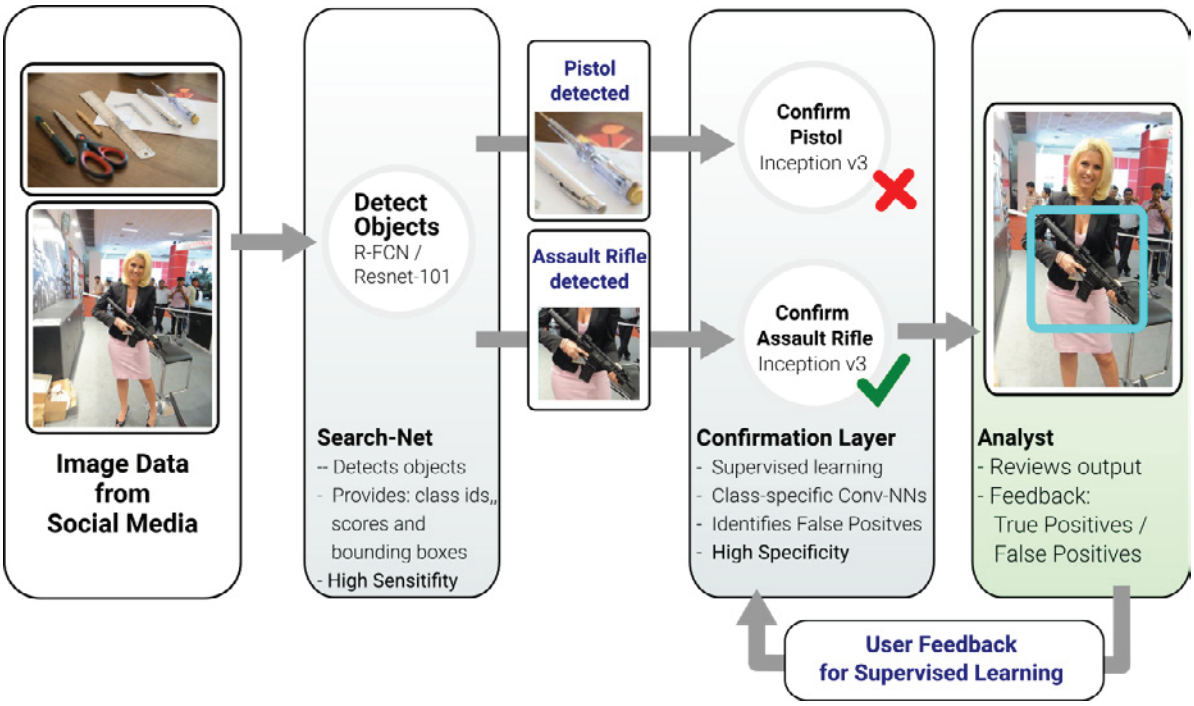
object detection and second, the Confirmation-Layer network which is used to revise the output of the first network and consists of multiple Inception v3 networks (Szegedy et al., 2015), one for each respective class of the Search-Net. The Search-Net analyzes the input data and returns class ids, scores and bounding boxes for the detected objects. The extracted objects are provided to the Confirmation Layer which evaluates whether a detection is correct (true positive) or incorrect (false positive) (see Fig. 1). The classifiers of the Confirmation Layer are continuously trained on the user feedback and thereby learn to detect systematic misclassifications of the Search-Net. The system was implemented using Python and the machine learning framework Google TensorFlow (Abadi et al., 2015), in particular its object detection functionality (Huang et al., 2017). All computations are conducted on an Intel i7 workstation equipped with a Nvidia Geforce GTX 1080 for GPU processing.

For all detected objects, the Search-Net and the classifiers of the Confirmation Layer calculate confidence scores based on the extracted features. By setting thresholds for both modules, the sensitivity and specificity of the whole system can be optimized for the specific use case.

2.2. Training Data

The Search-Net was trained on pictures collected from the ***Internet Movie Firearms Database*** (IMFD, 2017) for two classes (1162 pistols and 387 assault rifles) which were annotated manually. The classifier of the Confirmation Layer were trained on true and false positive samples for both classes. To generate the true positive data, the Search-Net was run on the training data and the image sections which contained the detected objects were extracted and validated manually (pistols: 1391 / assault-rifles: 734). Moreover, 5703 random images collected from ***Flickr*** (Flickr, 2017), which contained no weapons, were classified by the Search-Net. Here, 738 assault-rifles and 1211 pistols were mistakenly detected and were used to generate the false positive data. With this data, the Confirmation Net was trained to detect false positives of the Search-Net for both pistols and assault rifles separately.

Figure 1: 2-Pass Object Detector. The Search-Net provides class ids, scores and bounding boxes of detected objects. The Confirmation Layer checks the output for misclassifications. The analyst gives feedback regarding correct and incorrect classifications which is used for refinement of the Confirmation Layer network.



3. Results

The performance of the 2-Pass Object Detector was evaluated using a test data set with images collected from ***Flickr*** (Flickr, 2017) and ***Wikimedia Commons*** (Wikimedia, 2017), which consisted of 80 images depicting weapons (40 pistols / 40 assault rifles) and 734 random images with no weapons. The Search-Net was configured to classify an object as weapon if the confidence score exceeded 0.90. With this threshold, it identified 75 of the 80 images depicting weapons correctly, which corresponds to a true positive rate (sensitivity) of 93 %. However, 117 images were mistakenly classified to either show at least one pistol or assault rifle which corresponds to a true negative rate (specificity) of 85%. The output of Search-Net was then re-evaluated by the Confirmation Layer. Here, the threshold for the confidence score was set to 0.50. The Confirmation Layer was able to correct 75 false positive detections, but also reclassified 5 correctly identified weapons as false detection. Consequently, the specificity of the whole framework increased to 95% while the sensitivity was decreased to 87% (see Tab. 1). Some of the results are shown in Fig. 2.

Table 1: Results before and after reclassification by the Confirmation Layer

	True Positive	False Positives	Sensitivity	Specificity
Search-Net	75 of 80	117 of 734	93 %	85%
Confirmation Layer	70 (-5)	42 (-75)	87 %	95%

Figure 2: Results obtained from the test image data set. Top: All weapons have been correctly detected and annotated by the 2-pass object detector. Bottom left: All but one assault rifle have been correctly detected and annotated. Button right: Combination of two objects mistakenly classified as pistol.

4. Discussion

The 2-pass approach allows for a direct refinement of the object detector based on the user feedback. A common mistake of the Search-Net was for example to misinterpret a cell phone, which is held in a hand, as a gun due to shared geometrical features. After providing a few false positive samples, the Confirmation Layer was able to detect this systematic misclassification without the need to retrain the whole Search-Net.

At the same time, the 2-pass approach significantly reduces the time required for supervised learning. As discussed, analyst feedback is incorporated into the Confirmation Layer. The Confirmation Layer only processes defined parts of images and does not scan the whole source image as is required for the Search-Net. Hence, the Confirmation Layer can be re-trained to incorporate the user feedback within minutes. On the other hand, a complete training of the Search-Net will take, on the hardware used for this study, several hours to achieve good convergence of training results.

The 2-pass approach also allows to tune sensitivity and specificity depending on the requirements of the task at hand. For our test data set, we were able to increase the specificity from 85% to 95% without losing too much sensitivity (93 % to 87 %).

So what do these numbers mean for the daily workload of an analyst? Let's consider the following case: An analyst has obtained a data set with 10.000 images from a social media source and he wants to evaluate if members of that group pose with weapons on some of these photos in order to assess their threat potential, while it is not necessary to detect all occurrences of weapons. Let's assume that 10 of the 10.000 photos show a person with a weapon. Without technical support, the analysts would have to go through 10.000 / 10 = 1000 photos on average until he finds the first one showing a weapon. An automatic detector with a sensitivity of 93% and a specificity of 85% will find, on average, 9.3 pistols and 1499 false positives (15% out of 9990 possible false positives in 10.000 images) in the data. This means that (1499 + 9.3) / 9.3 ~ 162 images, on average, have to be checked to find the first picture showing a weapon. If the presented 2-pass object detector with a sensitivity of 87% and a specificity of 95% is used, this number once again decreases to (499 + 8.7) / 8.7, i.e. ~ 58 images. If the analyst checks 0.5 images per second, his workload for this specific analysis would have been reduced from initially about 33 minutes to about 2 minutes.

The presented object detection technique is not restricted to weapons but can be trained on any object. To scan for car plates and street signs can help to get evidence about the location where the photo has been taken and to identify group members. Scanning for symbols and logos can help to gain information about group affiliation.

We believe that this or a similar technique for image object recognition can greatly increase the efficiency of police work while potentially increasing privacy of users by limiting the amount of content explicitly monitored by police analysts.

Acknowledgements

This work was partially funded under the project "INTEGER" by the Federal Ministry of Education and Research, Germany, reference number FK 13N14377.

References

- Abadi, M. et al. (2015) TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems.
  Available from: http://download.tensorflow.org/paper/whitepaper2015.pdf
- Dai, J.; Li, Yi; He, K. & Sun, J. (2016) R-FCN: Object Detection via Region-based Fully Convolutional networks.
  Available from: ArXiv:1605.06409
- Flickr (2017),
  See: https://www.flickr.com
- He, K.; Zhang, X.; Ren, S. & Sun, J. (2016) Deep residual learning for image recognition. In The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 770-778.
- Huang, J.; Rathod, V.; Sun, C.; Zhu, M.; Korattikara, A.; Fathi, A.; Fischer, I; Wojna, Z.; Song, Y.; Guadarrama, S.; Murphy, K. (2017) Speed/accuracy trade-offs for modern convolutional object detectors.
  Available from: https://arxiv.org/abs/1611.10012
- Internet Movie Firearms Database (2017),
  See: http://www.imfdb.org/wiki/Main_Page
- Szegedy, C.; Vanhoucke, V.; Ioffe, S.; Shlens, J.; Wojna, Z. (2015) Rethinking the Inception Architecture for Computer Vision.
  Available from: ArXiv:1512.00567,
- Wikimedia (2017),
  See: https://commons.wikimedia.org/wiki/Main_Page

Licenses of images used

The following images from Wikimedia under a Creative Commons / attribution license were used in this paper. The use of these images does not mean the copyright holder endorses this work. Please see link for details of licenses.

1. https://commons.wikimedia.org/wiki/File:Anca_Verma_with_SIG_SAUER_rifle_at_DefExpo_India_2012.JPG
2. https://commons.wikimedia.org/wiki/File:USS_Mitscher_(DDG_57)_150129-N-RB546-055_(16433068885).jpg
3. https://commons.wikimedia.org/wiki/File:120516-A-WC501-101_(7325177330).jpg
4. https://commons.wikimedia.org/wiki/File:Flickr_-_Official_U.S._Navy_Imagery_-_Sailors_fire_M9_service_pistols_during_arms_qualifications..jpg

The following images from Flickr under a Creative Commons / attribution license were used in this paper. The use of these images does not mean the copyright holder endorses this work. Please see link for details of licenses.

1. Vesselin Dochkov, Tools, https://www.flickr.com/photos/vesselin/24270513429/
2. U.S. Army Europe, Training tools, ttps://www.flickr.com/photos/usarmyeurope_images/9266950263/