SAND2017-7563D

Exceptional service in the national interest

Sandia National Laboratories

Convolutional Neural Networks for Automatic Threat Detection in X-ray Images

Trevor Morris
University of California, Santa Barbara
M.S. Computer Science, June 2018

Tiffany Chien
University of California, Berkeley
B.A. Computer Science, May 2020

Sandia National Laboratories/NM, U.S Department of Energy
Manager: Judith Spomer; Mentor: Eric Goodman, Org. 9365
July 26, 2017

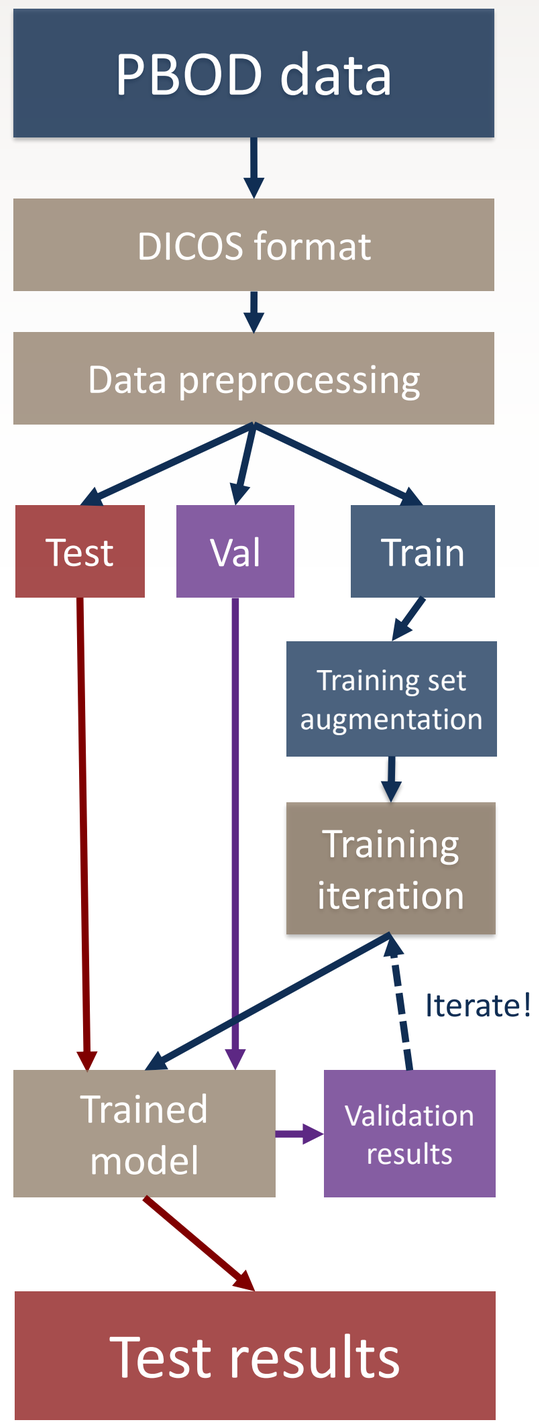
Abstract

This project applies Convolutional Neural Networks (CNNs) to the task of automatic threat detection, specifically explosives, in security X-ray scans of passenger baggage. Our data preparation methods preserve important features of the images while making our data compatible with different models. Using three different prebuilt state-of-the-art models and taking advantage of the properties of the X-ray scanner, we achieve reliable detection of threats. We employ transfer learning and training set augmentation to overcome our relatively small dataset. We also use visualizations to help interpret and adjust the training of our models.

Introduction

Convolutional neural networks (CNNs) have recently made huge gains in the image recognition space, made possible by large publicly available datasets like ImageNet, and more powerful computing systems. Since 2012, many new model architectures and principles have been developed, gradually improving performance as well as understanding of these models.

Our project is applying these models to X-ray baggage data. This is not currently a widely researched area, mostly because X-ray datasets are hard to come by. These images pose a significant challenge beyond standard images because of the amount of clutter there can be, and the limitations of X-ray in this setting. Clutter obviously makes threat detection difficult because objects can easily obscure other objects, and in X-ray, the "opacity" of different materials can make for many layers of obscuring clutter. One specific challenge in detecting explosive threats is that the model cannot use shape to identify threats, and can only use features analogous to "color" and "texture" (different materials have different X-ray response). The goal of our project is to explore the effectiveness of CNNs on this task, using data from the Passenger Baggage Object Database (PBOD).



U.S. DEPARTMENT OF ENERGY
National Nuclear Security Administration

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology and Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International, Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525. SAND No. 2011-XXXXP

Exceptional service in the national interest

Sandia National Laboratories

Data Properties

DICOS is an image format designed for security screening image data, adapted from the medical imaging format DICOM.

The scanners used to generate the PBOD data produce two channels in each image, analogous to RGB channels in a standard image. The first is intensity, measuring the magnitude of the x-ray response of the objects in the scan, which is what is often displayed as a grayscale image. The other channel is z-effective, which is an estimate of the "effective" atomic number of a material, based on its response and interaction with x-ray. We hypothesized that the z-effective channel would be very important to the detection of explosives. To get and use DICOS images, we converted from the proprietary format RCF, and then used an open-source DICOM reader.

Another feature of these scanning systems is that each scan generates projections from two different angles. Our model takes advantage of these two views by combining (average or max) the probabilistic predictions of the two views to generate one prediction for each scan.

Data Preparation

Before the models can use the images, they had to all be resized to the same size and square. Originally, the resizing was done by scaling, not preserving the aspect ratio. However, this could potentially stretch or skew important material properties, so we switched to a cropping approach that maintains the aspect ratio and the relevant part of the image. First, we trim off as much of the background as possible, by finding all pixels where the intensity value is above a certain threshold. This region is then padded on its shorter axis to make the image square, and then downsampled to the desired size.

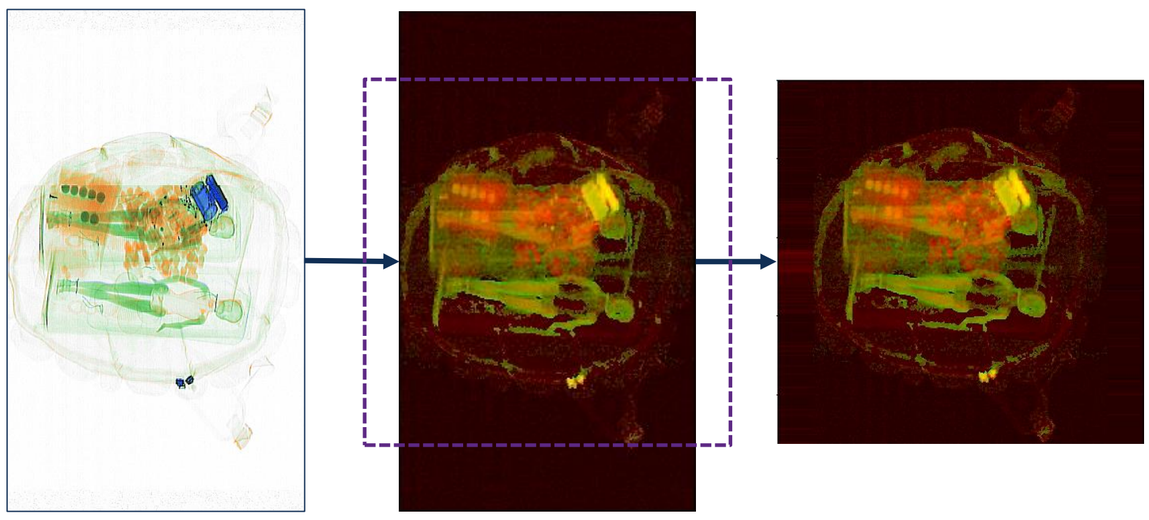


Figure 1. Each image is first converted to DICOS, with red representing the intensity channel and green the z-effective. Then, it is cropped to a square by cutting out the blank background and padding.

The CNNs we trained are huge, designed for training on 1.2 million ImageNet images in over 1000 categories. For this reason, it is very beneficial to modify our training set to augment its size. Using Keras' ImageDataGenerator class, during each epoch of training, every image (in the training set) is randomly flipped, rotated, shifted, and zoomed, so that after N epochs, the model will have been trained on N slightly modified training sets.

Because the two channels do not have the same magnitude and scale, the data is also normalized.

Visualizations

One common criticism of CNNs is that they are black boxes: the millions of trained parameters are uninterpretable to humans. However, it is possible to gain insight into how our models are identifying threats through visualizations.

One method is to use the network to synthesize images that will maximally activate a specific neuron. To do this, we start with an image of random noise, and apply gradient ascent to maximize the activation in the given neuron, similar to how gradient descent is applied during training to minimize error. This would typically produce unrecognizable, noisy images; we apply regularization to create more realistic images.

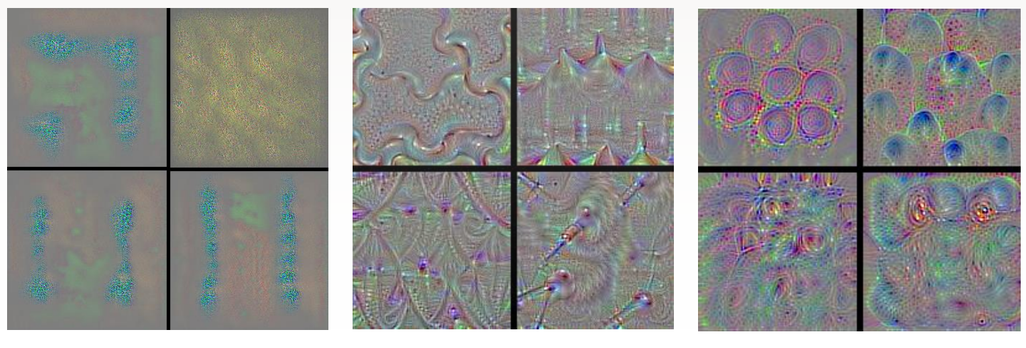


Figure 2. Constructed images that maximally activate a network's filters. From left to right, a network trained from scratch, a network pretrained on ImageNet, and then the pretrained network fine-tuned on our data.

The results help us understand how the network has adapted after being retrained, and gives us insight into the type of features it is looking for. In addition, by looking at neurons in different layers, we can see both higher-level and lower-level features. For example, we noticed that networks trained from scratch are looking more for patterns or texture, compared to pretrained, fine-tuned networks, which prompted us to look closer at the way we were processing our data.

Results

We ran three different prebuilt CNNs: Inceptionv3, Xception, and VGG19. These models are of varying design and size, with different priorities in efficiency and complexity. We also tried transfer learning with each model initialized to weights trained on the ImageNet dataset. The models were trained on about 6000 available PBOD images.

Model	Accuracy (%)	Training time (min)
Inceptionv3	85.28	49.73
Xception	86.32	88.85
VGG19	88.69	136.0
Inceptionv3 pretrained	87.23	75.67
Xception pretrained	87.63	91.62
VGG19 pretrained	90.60	86.20

Conclusions and Future Work

Overall, these are promising initial results, and CNNs are clearly a viable approach to the task of automatic threat recognition. In the future, model performance can be improved with more fine-tuning of parameters. In addition, PBOD has annotations to help train models to determine the precise location of a threat. This work could also be extended to 3D CT images.

U.S. DEPARTMENT OF ENERGY

National Nuclear Security Administration

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology and Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International, Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525. SAND No. 2011-XXXXP