X-Net: A Convolutional Neural Network for X-Ray Threat Detection

Ryan S. Park

Abstract— **This paper proposes X-Net, a novel deep learning architecture that enhances airport security through the detection of dangerous objects in X-ray luggage scans. Scanning of luggage is a critical part of aviation safety but is alarmingly unreliable due to human error. This endangers the safety of millions of airline passengers. To eliminate this human error, several deep learning concepts engineered for the analysis of X-ray baggage scans are introduced. The different concepts are all part of one model, called X-Net. X-Net employs a network of deep convolutional lateral stacks, which combine vertical residual transpose blocks with inter-layer connections. This combination allows for multi-directional gradient flow, resulting in richer and more robust internal feature representations. These innovations enable X-Net to perform exceptionally well on real-world baggage scans, significantly enhancing public safety and potentially saving thousands of lives: X-Net detects malicious items 400% more accurately and 4200% faster than a human TSA officer. Moreover, this proposed approach provides novel and empirically useful deep learning tools that strengthen other fields of computer vision.**

I. BACKGROUND

Now more than ever, airport safety is a high-priority national security issue. Two decades ago, 9/11 exposed the catastrophic results of lackluster aviation security; today, modern terrorism has made the task of protecting airports all the more significant. However, this task is not an easy one, as the ever-increasing rate of air travel has put a strain on security organizations like the TSA. Within the TSA's airport-related responsibilities, X-ray threat detection ("TD") is one of the most important, as it is the primary means of analyzing passengers' luggage. In this high-risk scenario, the TSA has been alarmingly ineffectual. Reports by *Forbes*, *ABC News*, *Newsweek*, and *The New York Times* indicate that the TSA has failed to catch 70-95% of dangerous items in multiple undercover government tests [2], [3], [5], [8]. This astronomical failure rate is no surprise. Ultimately, the failure of baggage screening tests is a result of human error. The logical solution is to remove the human component, i.e., employ automatic threat detection ("ATD") to decrease the failure rate. This paper explores a deep learning model, named X-Net, as an accurate and efficient ATD system to address the issue of X-ray TD.

II. X-NET DESIGN

X-Net combines the YOLO detection head from Redmon et al. (2018) with a novel backbone ("X-Net-backbone"). The detection head is a three-tier residual convolutional network, which detects objects at different input image resolutions [7]. The backbone is the feature extractor, which condenses the information in the input image such that the detection head can then detect objects. This research introduces "lateral stacks" as the core of X-Net-backbone (inspired by Lin et. al (2017)'s research on feature-pyramid networks). The X-Net-backbone has a main horizontal branch along with several lateral stacks (mini neural

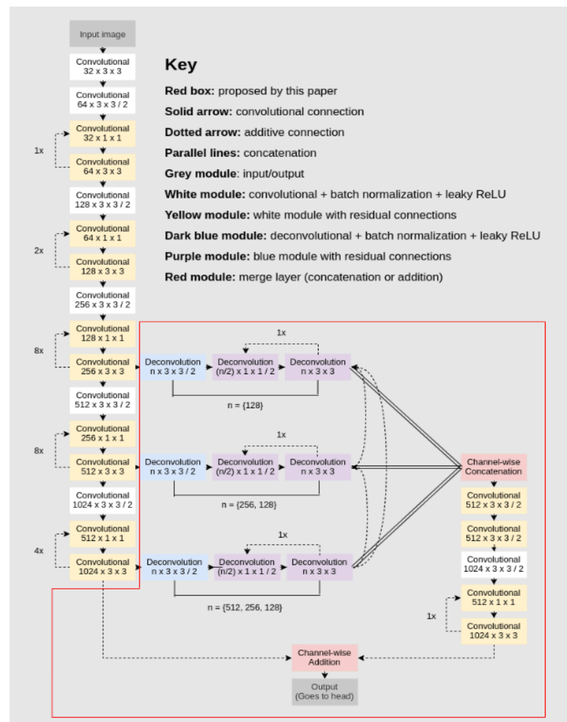

Figure 1. X-Net-backbone. *Deep lateral stacks are purple/blue and are condensed for viewing purposes. Because of space constraints, "vertical" or "lateral" layers appear horizontal, and vice versa.*

networks that stem from the main convolutional branch). The outputs of these stacks are concatenated channel-wise and serve as the input to another network, called the secondary horizontal branch. The output of this branch is then added with the output of the main horizontal branch to form the final output of the X-Net-backbone. See Fig. 1 for a visualization of this architecture, which streamlines information flow and prevents gradient vanishing deep in the network.

III. METHODOLOGY

X-Net is trained and evaluated on SIXray [6]: an X-ray baggage scan dataset created by the Pattern Recognition and Intelligent System Development Laboratory of the University of Chinese Academy of Sciences. SIXray contains 1,059,231 X-ray baggage scans with 12,245 malicious objects, including knives, scissors, guns, wrenches, and pliers. In total, X-Net is trained on ~9,000 X-ray scans with MS-COCO-pretrained weights initialization in early layers. Testing and validation sets were held-out using ~500 scans each, with no overlap between train/validation/test splits.

X-Net is programmed using Keras and TensorFlow. Open-source code [1] is used for evaluation and results. Code from this study is publicly available at https://github.com/orangese/x-net.

Training utilizes a GeForce RTX 2060 for 72 hours. Due to hardware limitations, a resource-conscience training schema is employed. X-Net is trained in three different stages, using Adam with a batch size of 1 and learning rates = [0.001, 0.0001, 0.00001]. This annealed training strategy initially encourages exploration of parameter space (learning rate = 0.001) and then encourages convergence towards a minimum of the loss function (learning rate = 0.00001).

IV. RESULTS AND DISCUSSION

The main innovation of X-Net are the aforementioned "deep lateral stacks": these are mini-networks that branch off of several points within the convolutional backbone of X-Net. Unlike in previous research, which used largely linear network structures like the Inception and ResNet families [6], X-Net incorporates information across several resolution "checkpoints" along the backbone. Thus, the detection of smaller malicious objects is emphasized compared to previous convolutional networks that lossily compress tensor dimensions deep into the network.

Two evaluation metrics from [6] are used to evaluate X-Net: classification accuracy (measures detection of malicious items) and localization accuracy (measures the ability to identify the location of malicious items). In this section, accuracy is computed using mean average precision ("mAP"), a standard computer vision metric used to assess object detection algorithms. All reported percent results are relative percent increases, not absolute increases in mAP.

With respect to classification accuracy, X-Net outperforms the next-best ATD model by 9.93% and plain YOLOv3 by 11.03% (Table 1):

TABLE I. CLASSIFICATION MAP (%) OVER ALL CLASSES AND MODELS

Method	Item Category					
	Gun	*Knife*	*Wrench*	*Pliers*	*Scissors*	*Overall*
ResNet-101 [6]	87.65	84.26	69.33	85.29	60.39	*77.38*
ResNet-101+CHR [6]	85.45	87.21	71.23	**88.28**	64.68	*79.37*
Inception-v3 [6]	90.05	83.80	68.11	84.45	58.66	*77.01*
Inception-v3+CHR [6]	88.90	**87.23**	69.47	86.37	65.50	*79.49*
YOLOv3	94.65	77.75	71.03	83.10	66.96	*78.70*
X-Net	**96.28**	83.52	**78.71**	88.16	**90.24**	***87.38***

More importantly, X-Net decreases human error by several orders of magnitude, achieving a 399.31% gain in classification mAP and a 4247.83% gain in speed over a TSA officer (Fig. 2):

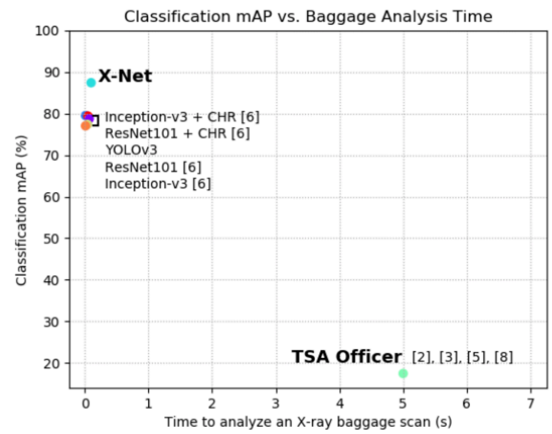

Figure 2. Classification mAP (%) vs. Baggage Analysis Time (s).

With respect to localization accuracy, X-Net improves overall mAP by 17.93% (Table 2):

TABLE II. LOCALIZATION MAP (%) OVER ALL CLASSES AND MODELS

Method	Item Category					
	Gun	*Knife*	*Wrench*	*Pliers*	*Scissors*	*Overall*
ResNet-101 [6]	73.77	65.13	28.34	62.24	21.02	*50.10*
ResNet-101+CHR [6]	80.86	73.85	52.41	9.30	40.34	*51.35*
Inception-v3 [6]	79.94	**75.38**	59.36	59.58	40.34	*62.92*
Inception-v3+CHR [6]	78.70	74.36	52.41	59.96	52.27	*63.54*
YOLOv3	78.35	55.11	43.68	51.90	39.39	*53.68*
X-Net	**87.95**	72.01	**67.24**	**73.90**	**73.55**	***74.93***

V. CONCLUSION

Ultimately, X-Net advances both deep learning theory and its applications. First, it provides empirical justification for the proposed lateral stacks. Specifically, X-Net increases detection and localization of the smallest class (scissors) by 34.77% and 40.71%, implying that lateral stacks are especially useful for detecting hard-to-see objects. Similar accuracy gains can be noted for the other objects in Tables 1 and 2. However, the accuracy gains for larger objects (guns, knives, wrenches, etc) were smaller. This result is consistent with the intended design of X-Net: unlike previous networks, X-Net is able to access information at various stages within the network (i.e., both high-resolution and low-resolution) due to its deep lateral stack paradigm. As a result, it can detect smaller objects with relative ease. Increasing detection accuracy for the larger objects, however, is a potential extension of this research that will likely be addressed as convolutional network design continues to evolve.

X-Net outperforms the TSA by 399.31% in terms of malicious item detection accuracy and is incredibly fast, achieving a 4247.83% reduction in time per scan. Therefore, X-Net is a critical enhancement of airport security and a practical solution to the problem of X-ray TD. If integrated into security systems, it would detect many threats and save many lives, thereby helping to preserve public safety

VI. REFERENCES

[1] Cartucho. (2018, March 18). Retrieved from https://github.com/Cartucho/mA
[2] Goldstein, M. (2017, November 9). TSA Misses 70% Of Fake Weapons But That's An Improvement. *Forbes*. Retrieved from

https://www.forbes.com/sites/michaelgoldstein/2017/11/09/tsa-misses- 70-of-fake-weapons-but-thats-an-improvement/#1d42a8b32a38
[3] Kerley, D., & Cook, J. (2017, November 9). TSA fails most tests in latest undercover operation at US airports. *ABC News*. Retrieved from https://abcnews.go.com/US/tsa-fails-tests-latest-undercover-operation-us-airports/story?id=51022188
[4] Lin, T.-Y., Dollar, P., Girshick, R., He, K., Hariharan, B., & Belongie, S. (2017). Feature Pyramid Networks for Object Detection. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). doi: 10.1109/cvpr.2017.106
[5] Meza, S. (2017, October 11). TSA Fails to Spot Weapons More than Half the Time. *Newsweek*. Retrieved from https://www.newsweek.com/tsa-failshalf-time-706568
[6] Miao, C., Xi, L., Wan, F., Su, C., Liu, H., Jiao, J., & Ye, Q. (2019). SIXray: A Large-scale Security Inspection X-ray Benchmark for Prohibited Item Discovery in Overlapping Images. Retrieved from https://arxiv.org/pdf/1901.00303.pdf
[7] Redmon, J., & Farhadi, A. (2018). YOLOv3: An Incremental Improvement. Retrieved from https://pjreddie.com/media/files/papers/YOLOv3.pdf
[8] Smith, J. F. (2015, June 2). Head of T.S.A. Out After Tests Reveal Flaws. *The New York Times*. Retrieved from https://www.nytimes.com/2015/06/02/us/head-of-tsa-out-after-tests-reveal-flaws.html