Briefing Note BN22/02. Date of publication: March 21, 2022.

Metior Telum: Measure the Weapon

Janis Dalins and Campbell Wilson

AiLECS Lab, Melbourne, Australia
Email: janis.dalins@afp.gov.au,janis.dalins@monash.edu; campbell.wilson@monash.edu

The main challenge of machine learning (and indeed any automated analysis) is understanding and representing what an algorithm will actually encounter once deployed. Identifying representative data is one challenge, gathering it is another. When it comes to media, a quick, effective approach heavily utilised throughout industry and academia is the web scrape. The AiLECS Lab's VALID project[1] explores the inherent dangers of such data gathering, both from ethical and technical perspectives. For the sake of brevity, we summarise the main problems thus:

1. Dubious permissions. It is rarely, if ever, known if the original imagery holders consent to their images being used for the research, or even to have been hosted online in the first place; and
2. Quality Assurance. In the absence of expert knowledge, the quality and representative nature of the data being gathered is unknown. Is there a good spread of angles, lighting, backgrounds? If you're seeking to detect specific objects, are they even labelled? Image annotation is a slow, mundane process, often leading to questionable outputs.

Whereas Item 2 is a common issue in machine learning, Item 1 is particularly problematic from a law enforcement perspective. Perceived overreaches into user privacy undermine trust community trust, damaging the social contract inherent to policing by consent.

1. The Problem

Commercially available object detection services were observed to underperform in recent AFP investigations, both in terms of accuracy but above all granularity - labels such as 'gun', 'rifle', 'shotgun' and 'pistol' were regularly misclassified, with rifle and shotgun in particular often confused[2]. Whereas not a major concern in and of itself (the presence of a firearm being the main priority), the inability to infer a difference between, say, a .22 calibre bolt action rifle and an AK-47 limits opportunities for triage.

Attempts to train on richer data immediately ran into limitations - open source firearm datasets tend to feature well lit, well positioned images such as Figure 1, with minimal examples of weapons being held or fired.

Figure 1. Wikipedia Commons image for M16A2 - taken from https://commons.wikimedia.org/w/index.php?curid=80178500

[1] https://ailecs.org/project/the-valid-project/

[2] This confusion was also observed within open source datasets such as Google Open Images, perhaps indicating a lack of domain knowledge amongst labellers

Specialised sites such as online firearms marketplaces and Reddit threads contain more representative imagery, but this raises legitimate questions of consent and perceptions of unfair treatment in the collection of this data. That is, while the work is done with the best of intentions, could it also be perceived as mass surveillance of activities undertaken by a social group, being lawful firearm owners?

2. The Solution

In the absence of any representative and ethically guaranteed sources, we decided to make our own. Synthetic datasets are an established means for overcoming a lack of data - i.e. if you need more, make what best appears to you as a representative sample. Helpfully, the Australian Federal Police's (AFP) Firearm Reference Library contains well over 8,000 exemplars, ranging from pocket pistols to heavy machine guns, so access to 'representative' items was assured.

2.1 Capture

Digitisation of a representative sample is the next challenge. A scan of 600 items from the reference library is currently underway, conducted by a photogrammetry and scanning vendor with a strong history in CGI for movies and gaming - tasks strongly paralleling the problem at hand. The photogrammetry is being undertaken using a customised rig, guaranteeing consistency across items through the use of 64 statically mounted SLR cameras and controlled lighting conditions. Once captured, the scan and photogrammetry are combined to form a digital asset.

2.2 Processing

A happy side-effect of creating one's own digital assets is the ability to *render*[3] them in any way one desires - one scanned item being capable of rendering using infinite combinations of angle, distance, lighting and background.

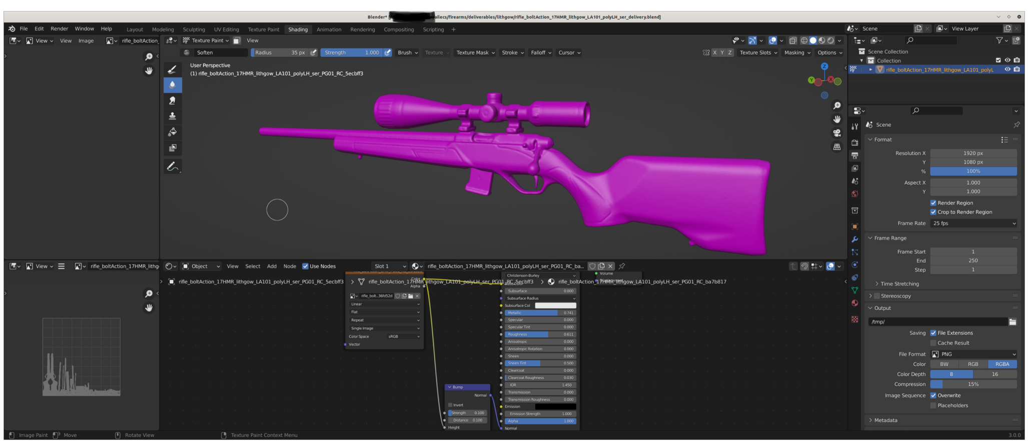

Figure 2. Sample screenshot of rendering process

Current processing iterations are assigned using random combinations of variables such as lighting, but it is reasonable to assume better performance will be achieved through weighting towards more commonly seen scenarios, though the extent of such improvement is yet to be measured.

It should be noted the image shown in 3 is **not** the highest quality available in processing. Most (if not all) algorithms currently in use for image processing require some degree of downsampling. It was observed in early testing that photorealistic renders did not significantly improve performance, hence for cost reasons a 'medium' quality render was specified. Given the capture of data at high precision, this can be revisited in future if requirements sufficiently change.

2.3 Deployment

Automation is key for the project to support large (100,000+ image) on-demand datasets, complete with bounding boxes[4]. Given we've set the locations of our objects within candidate images, we can automatically assign bounding boxes, and push the images direct into our training environment - for our trial, as a dataset within Google Cloud.

[3] In this case, presenting the 3 dimensional model as part of a 2 dimensional scene

[4] Shapes denoting the presence of an object on an image, not dissimilar to what appears in Figure 4.

Figure 3. Sample finished render (image courtesy Wysiwyg 3D)

From there, development, testing and deployment of any models becomes a standard matter of data science and applications. In our case, we utilised AutoML as a quick (in terms of user hours) means for establishing whether the datasets retain enough discernable features for effective, representative training - in this case, for object detection.

Figure 4 shows the model's performance when presented with a candidate rifle on a non-typical angle - it correctly recognises a Lithgow LA-101 rifle, in this case, with different sights than the sample shown in Figure 2. This particular model was trained in 24 hours using a small (approx 2000 images per class) dataset of 5 distinct weapons, and has been observed to underperform. We anticipate this particular scenario to improve rapidly as the dataset size increases.

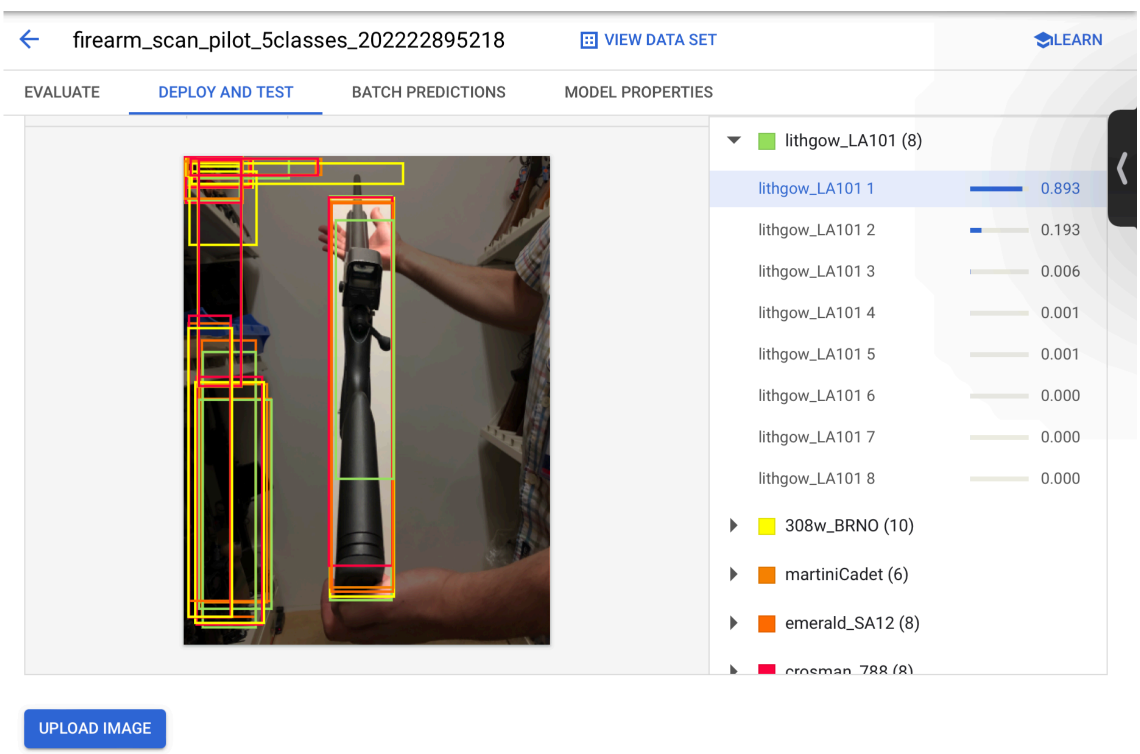

Figure 4. Pilot detector result - Lithgow LA101 rifle

2.4 Conclusion

Metior Telum's core purpose is to test and prove the feasibility of digitised firearm libraries for use in computer vision based tasks. Early results are promising, with an automated pipeline now taking shape from initial render right through to model monitoring and tuning.