Instance Segmentation of Armoured Fighting Vehicles using Fully Convolutional Object Detection and Domain Randomisation

Ville Rissanen
Department of Military Technology
National Defence University
Helsinki, Finland
ville.rissanen@linux.com

Einar Eidstø
Faculty of Medicine
University of Helsinki
Helsinki, Finland
einar.eidsto@gmail.com

Pyry Virtanen
Department of Applied Physics
Aalto University
Helsinki, Finland
pyry.virtanen20@gmail.com

Eemil Praks
Department of Applied Physics
Aalto University
Helsinki, Finland
eemil.praks@gmail.com

Tenho Korhonen
Department of Computer Science
Aalto University
Helsinki, Finland
korhonentenho@gmail.com

Emil Toivonen
Faculty of Science and Engineering
Åbo Akademi University
Turku, Finland
toivonen.ebe@gmail.com

Jouko Vankka
Department of Military Technology
National Defence University
Helsinki, Finland
jouko.vankka@mil.fi

Keywords—instance segmentation, object detection, armoured fighting vehicle

I. Introduction

Images of real modern military equipment are limited in availability and quantity which makes acquiring enough images to train a reliable neural net a challenge. However, research in using synthetic images as training material has yielded positive results even when real images are abundant [1]. This study researches the production and effect of using synthetic training data for use in a military context where real training data is not available in excess. We have developed a deep learning model prototype for detecting and segmenting armoured fighting vehicles (AFV) in different environments from images and video using CenterMask2 neural net architecture [2].

II. Method

A. Related Work

There is much research into object detection and instance segmentation of AFVs using various image sources [3,4,5,6], but we found no research that utilised synthetic images (domain randomisation) in the training of the models.

B. Technology

We chose CenterMask2 for the tasks of AFV detection and segmentation due to its state-of-the-art precision and ability forgo some precision for evaluation performance applicable for live video. The backbone used was deformable VoVNet2 [7]. We trained the model using real images of AFVs and synthetic images produced with a 3D modelling software. Additionally, data augmentation was applied to the images in the form of flipping the images and resizing the images to four different resolutions (640x480, 640x360, 520x576, and 600x 600).

C. Real image collection and curation

Real images were collected from the Internet automatically and then manually reviewed. The review process consisted of simply removing images that were not real AFVs thus eliminating false positives, various art pieces, and images of toys. Over 5500 unique curated images were collected with AFVs in various combat and non-combat situations and environments.

D. Rendering of synthetic images

Synthetic images were produced using the 3D modelling software Blender [8] with its embedded Python scripting environment to automate the process of rendering randomised images of AFVs. Randomised elements in the synthetic images include colour, angle of the source and intensity of the ambient light, background environment, “flying distractors” [1], camouflage pattern, 3D translation, and rotation of the AFV. Images of the AFVs were also rendered at varying angles and distances. The angle of the viewport was constrained to a hemisphere around the top half of the vehicle.

Fig. 1. An example of a synthetic training image.

The ambient light was simulated using a light source that's virtually infinitely far away from the scene of the viewport. The colour of the ambient light varied in a spectrum between white, yellow, orange, and red. The angle of the ambient light was constrained to a hemisphere around the top half of the AFV. The intensity of the simulated ambient light varied from dim moonlight (0.05 lux) to bright sunlight (100 kilolux).

The background environment was produced from a set of terrain ranging from desert to urban to forest at different times of the day and weather conditions. Background images were checked to not contain any AFVs in the picture.

Flying distractors are basic 3D geometries included in the rendered image that are shown to encourage more complex behaviour in neural nets during training [1]. A ground-level polygon mesh was used to partially cover the traction elements.

Camouflage patterns were applied to the AFVs and flying distractors were selected from a pool of modern AFV camouflage patterns available publically, few historical patterns, and monocoloured. Roughly appropriate camouflage for the background terrain and weather conditions were selected.

The 3D translation and rotation of the AFV are randomised, but it was ensured a majority of the vehicle remained within the viewport. An example image produced is featured in Figure 1.

E. Preparing images for training

The real and synthetic AFVs in the images were manually assigned bounding boxes and polygon segmentations as annotations for the training of the deep learning model. The model was trained from scratch on the dataset with a single Graphics Processing Unit (GPU) for a week. 20% of the all images were withheld for testing. The remaining set was split into 80:20 train:validation at random while ensuring only real images were selected for the validation set. In other words, for the final test dataset only real images have been used.

III. Results

The model can detect AFVs in plain sight with very high accuracy and partly obstructed or camouflaged tanks with good to high accuracy.

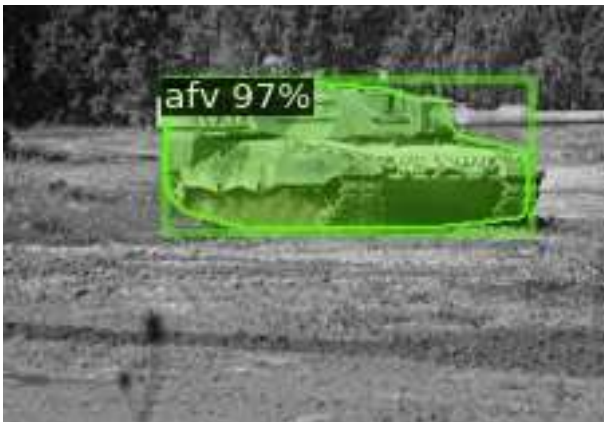

Fig. 2. An example of a succssfully detected and segmented Leopard main battle tank. Picture courtesy of Jarno Kovamäki, Finnish Defence Forces.

The average precision of the model with and without synthetic images added to the dataset is presented in Table I. Adding synthetic images to the training data yields a model with about 5% higher overall accuracy.

TABLE I. Model Training Results

Model	Results		
	Mask AP [a] (IoU [b]) With synthetic images	***Mask AP [a] (IoU [b]) Without synthetic images***	***Instance Size [c]***
CenterMask2	0.399 (0.5:0.95)	0.364 (0.5:0.95)	all
CenterMask2	0.449 (0.75)	0.391 (0.75)	all
CenterMask2	0.543 (0.5)	0.512 (0.5)	all
CenterMask2	0.132 (0.5:0.95)	0.113 (0.5:0.95)	small
CenterMask2	0.358 (0.5:0.95)	0.332 (0.5:0.95)	medium
CenterMask2	0.415 (0.5:0.95)	0.394 (0.5:0.95)	large

[a.] Average Precision

[b.] Instance over Union

[c.] Small instance size is less than 32 pixels and large is over 96 pixels with medium in between..

[d.] Results based on COCO dataset evaluation standards. (https://cocodataset.org/#detection-eval)

Fig. 3. Left: original image presented to the model. Center: Parts of the image that activate the model least are blurred. Right: A heatmap of the pixels that activate the network the most. Picture courtesy of Jarno Kovamäki, Finnish Defence Forces.

In addition to the ability to detect and segment instances of AFVs in images, the model could be used to improve camouflage technology by inspecting which elements of the detected AFVs contribute the most to successful detection. This can be achieved by feature inversion of the model as seen in Figure 3 [5].

REFERENCES

[1] J. Tremblay, A. Prak1ash, D. Acuna, M. Brophy, V. Jampani, C. Anil, T. To, E. Cameracci, S. Boochoon and S. Birchfield, "Training Deep Networks with Synthetic Data: Bridging the Reality Gap by Domain Randomization," in CVPR 2018 Workshop on Autonomous Driving, 2018.

[2] L. Youngwan and P. Jongyoul, "CenterMask: Real-Time Anchor-Free Instance Segmentation,", in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2020.

[3] Briggs, Ralph & Goldberg, Joseph. (1995). "Battlefield Recognition of Armored Vehicles." in Human Factors The Journal of the Human Factors and Ergonomics Society,37(3), pg.596-610 · September 1995

[4] Leslie M. Novak, Gregory J. Owirka, William S. Brower and Alison L. Weaver. "The Automatic Target-Recognition System in SAIP" in The Lincoln Laboratory Journal,10(2), 1997

[5] Steven K. Rogers, John M. Colombi, Curtis E. Martin, James C. Gainey, Ken H. Fielding, Tom J. Burns, Dennis W. Ruck, Matthew Kabrisky, Mark Oxley. "Neural networks for automatic target recognition" in Neural Networks,8(7-8), pg.1153-1184, 1995

[6] X. Xiaozhu and H. Cheng. "Object Detection of Armored Vehicles Based on Deep Learning in Battlefield Environment," in 4th International Conference on Information Science and Control Engineering (ICISCE), Changsha, pg.1568-1570, 2017

[7] L. Youngwan, H. Joong-won, L. Sangrok, B. Yuseok and P. Jongyoul, "An Energy and GPU-Computation Efficient Backbone Network for Real-Time Object Detection" in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2019.

[8] [Software] B. O. Community, "Blender – a 3D modelling and rendering package," Stichting Blender Foundation, Amsterdam.

[9] M. Du, N. Liu, Q. Song and X. Hu, "Towards Explanation of DNN-based Prediction with Guided Feature Inversion," in KDD2018, 2018.