Identifying Maritime Vessels at Multiple Levels of Descriptions using Deep Features

David Bo Heyse, Nicholas Warren, and Jelena Tešić

Computer Science Department, Texas State University, San Marcos TX

ABSTRACT

Varying resolution quality of operational data, size of targets, view occlusions, and large variation in sensors due to nature of overhead systems as compared to consumer devices contribute to degradation of the maritime vessel identification. We exploit the maritime domain characteristics to optimize and refine the deep learning Mask-RCNN framework for training generic maritime vessel classes. Maritime domain, compared to consumer domain, lack alternative targets that would be incorrectly associated as maritime vehicles: this allows us to relax the parameter constraints learned on urban natural scenes in consumer photos, adjust parameters of the model inference, and achieve robust performance and high AP measure for transfer learning scenarios. In this paper, we build upon this robust localization work, and extend our transfer learning work to new domains and datasets. We propose new approach for identifying specific category of maritime vessels and build a refined multi-label classifier that is based on deep Mask-RCNN features. The classifier is designed to be robust to domain transfer (e.g. different overhead maritime video feed), and to the noise in the data annotation (e.g. vessel is not correctly marked or label is ambiguous). We demonstrate superior category classification results of this low shot learning approach on publicly available MarDCT dataset.

Keywords: Deep neural network, Machine learning, Computer Vision, Data Science, Transfer Learning

1. INTRODUCTION AND RELATED WORK

In the past few years, systems for classifying and segmenting objects in images and videos established an expected performance baseline in consumer domain. Deep Convolutional Neural Network based systems perform various computer vision tasks such as image classification, object detection, semantic-segmentation, human body joint localization, and face recognition on consumer curated examples with super accuracy.[1–6] Complex Machine Learning algorithms, such as Deep Neural Networks (DNN), require a significant number of labeled training samples to perform well due to the large number of network parameters that need to be trained. Pascal VOC,[7] ImageNet,[8]and COCO[9] benchmarks motivated the breakthroughs in the field as training samples were collected through well executed and expensive crowd sourcing endeavor to label millions of object instances in imagery created by consumers using their hand-held devices.

To achieve similar performance in other domains, one has to consider the replication of similar process at comparable scale, and that is prohibitive in the domains of news, agriculture, archived cultural data, climate science, medical science, astronomy, space, underwater exploration, aerial imagery, satellite imagery, underwater imagery, and drone-captured imagery - there simply exist no crowd sourcing effort or labeling uniformity to achieve comparable benchmark at such a large scale. Transfer learning research problem focuses on storing knowledge gained while solving one problem and applying it to a different but related problem, and it has gained traction in use for domain adaption problem in computer vision. In domain adaption problem, we focus on utilizing multiple existing source data to build a model that will perform well on different but related dataset. For tasks where a sufficient number of training samples is not available, a DCNN trained on a large dataset for a different task is tuned to the current task by making necessary modifications to the network and retraining it with the available data.[10–13] Lately, multiple groups proposed an one shot learning approach for deep learning setup, and showed it to be consistent with normal methods for training deep networks on large data.[10,14] Domain adaptation of DNN systems has been used to produce segmentation maps and to improve category identification when applied to satellite imagery and remote sensing.[15–17]

Send correspondence to Dr. Jelena Tešić: E-mail: jtesic@txstate.edu.

NAVAIR Public Release 2019-349. Distribution Statement A – “Approved for public release; distribution is unlimited"

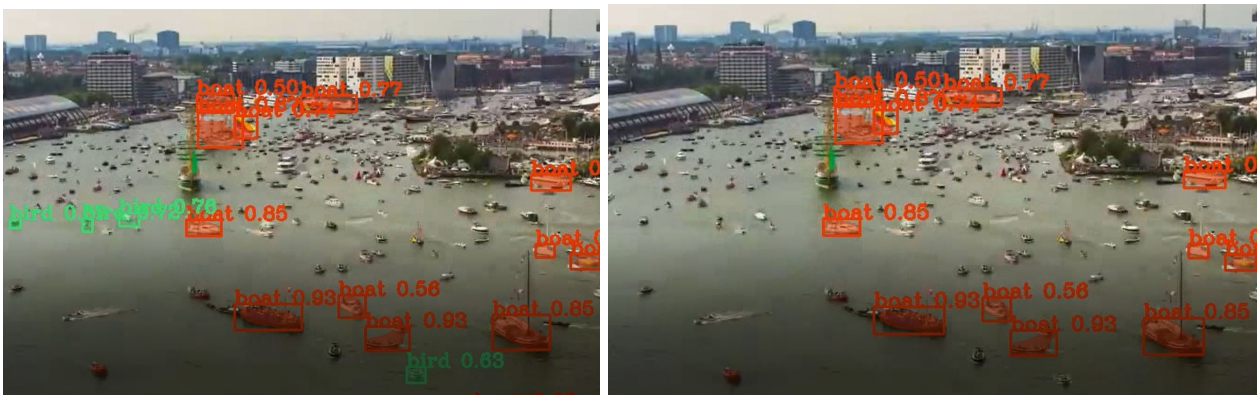

Figure 1. State-of-the-art in Deep Convolutional Neural Networks (DCNN) is consumer data. Sail 2015 Amsterdam video on YouTube[18] exhibits slight change in operating conditions (overhead camera, small objects as in this YouTube video). (Left) This change causes the system to identify small boats as birds. (Right) Training with other available maritime data improves classification precision, not recall.

In this paper we focus on object detection and recognition at multiple levels of description in maritime overhead imagery. In our previous work we have demonstrated the successful single source domain adaptation from consumer and maritime data sources to maritime object recognition.[19] In this paper, we focus on three different tasks: (a) multi-source domain adaptation of baseline models; (b) identification of maritime vessels when there is no maturity of annotated corpus, and (c) analysis of semantically relevant objects with large support from visual ontology standpoint in maritime domain.[20]

2. MULTI-SOURCE DOMAIN ADAPTATION

Low resolution quality of operational data, size of objects of interest, view occlusions, and crowded scenes degrade the performance of state-of-the-art DCNN when applied to overhead sensor and shipboard data. For Maritime datasets, the best algorithms struggle with objects that are small (distant objects) or with the distorted view (sun glare), which are common problems in ocean environments. Humans have no issues in recognizing objects in videos with similar conditions, but state-of-the-art machine learning algorithms break when there is a slight change in the operational environment. Figure 1 illustrates how the state-of-the-art DCNN model[4] produces different labels and candidate objects based on the training dataset. Deep neural networks trained on large corpora of labeled consumer images provide a robust generalized modeling start, and initializing a network with transferred features from almost any number of layers produces a boost to generalization.[21] In our work, we rely on this finding and expand from a consumer dataset to maritime domain, and adapting the deep neural networks system and parameters to reflect the target domain.

Intersection over Union (IoU) is an evaluation measure for the accuracy of object localization in an image, see PASCAL VOC,[7] ImageNet,[8] and COCO[9] benchmarks. IoU takes the set A of proposed object pixels within the proposed bounding box by the detector and the set of true object pixels B and calculates: $IoU(A, B) = A \cap B \div A \cup B$. In consumer benchmarks the detector performance is a hit if IoU of proposed detection A and ground truth B is larger than a threshold, otherwise it was a fail. In our previous work, we evaluate the performance of detectors using different measures of IoU, and evaluate performance sensitivity for maritime domain.[19] Our finding was that the change of IoU does take into consideration the influence the measure has on the small objects, and our recommended setup is to lower IoU for maritime overhead datasets. In this paper we adopt IoU threshold to be 0.5 for all experiments.

Model Refinement in Dynamic Scenarios allows user to fine tune models based on the domain requirement. As illustrated in Figure 2, data distribution can very from application to application, and maritime domain annotations are sparse. Our goal is to save and re-use any existing labeling in consistent manner. We reuse existing Deep Neural Network systems and fine-tune them to new datasets and new labels. Please see the extended discussion and mitigation of adversarial data influence in Section 3.

NAVAIR Public Release 2019-349. Distribution Statement A – “Approved for public release; distribution is unlimited"

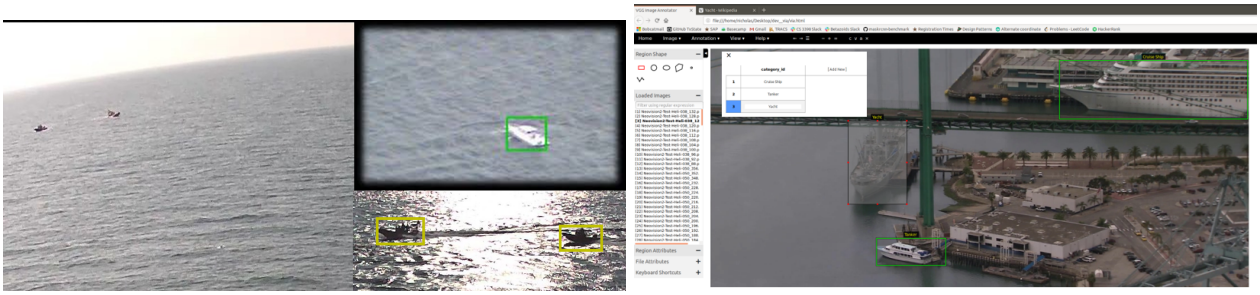

Figure 2. (Left) Examples of the multi source maritime dataset: angle, glare, and object size vary in overhead imagery; (Right) AIM VIA extension to support adding multiple labels to the same region.[22]

3. TRAINING DATA

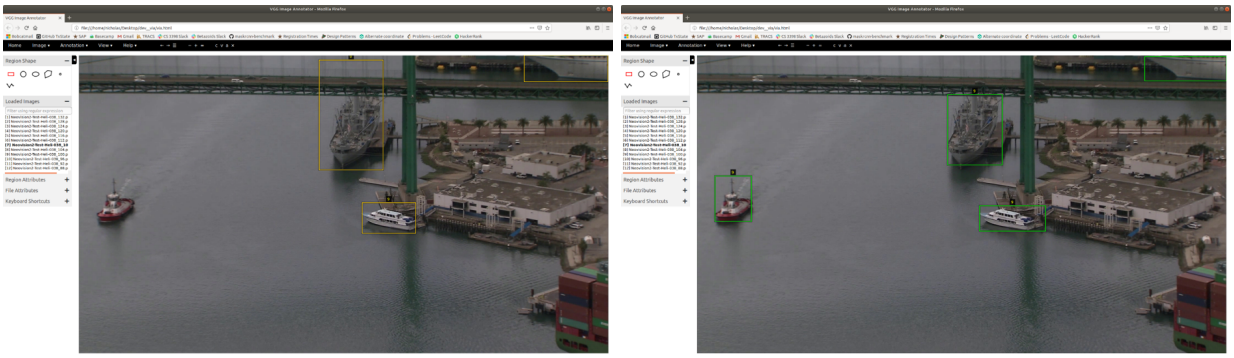

Figure 3. VIA tool adaptation for asset identification and monitoring: (Left) image shows automatic annotations (either ground truth or model inference). The system has missed several instances of boat. (Right) Analyst adds new annotations draws missing bounding boxes.[22]

A Deep Neural Network framework is inherently static, and training is mostly done in one location offline and the model is utilized for mass consummation e.g. image-to-text, identify consumer object in cell phone images or recognize a face. High confidence of the trained model is ensured by (a) a high number of training data and context filtering, and (b) lack of adversarial examples. In this section, we address both of these requirements in the context of maritime applications. Typical reconnaissance mission using maritime overhead imagery includes multiple EO sensor feeds, and a group of analysts that are reviewing these feeds from different location and perspective. Deep Convolutional Neural Networks for surveillance and monitoring needs to be utilized in a more dynamic environment: the sensor feeds have greater variance than consumer images. There is less labeled data available and the application of machine learning models for asset localization and identification varies due to the different objectives e.g. localizing, identifying, monitoring assets, or generating alerts. Deep Neural Networks are vulnerable to adversarial attacks in the form of subtle perturbations to inputs that lead a model to predict incorrect outputs. For images, such perturbations are often too small to be perceptible, yet they completely fool the deep learning models.[23] While the perceived scenario is free of purposeful adversarial annotations, the simple human error can have the same effect on the system, and it is more emphasized in the refinement scenario when the number of examples is small and one bad annotation in the small finite set can cause the error surface to have a strongly sub-optimal local minimum.[24]

We have adapted VIA VGG Image Annotation Tool[25] for analyst to foster this collaboration and persistent target labeling and intermittent modeling for real-time asset monitoring, and re-use existing annotations. Asset Identification and Monitoring for VIA provides analysts with an interface to (1) identify new objects of interests in maritime video feeds, (2) correct existing annotations, either from previous analyst or result of model inference, and (3) add additional labels for multiple levels of description. An analyst spots an asset of interest in a frame that has not been labeled by the DNN system, and using the AIM annotation functionality, then localizes and

NAVAIR Public Release 2019-349. Distribution Statement A – “Approved for public release; distribution is unlimited"

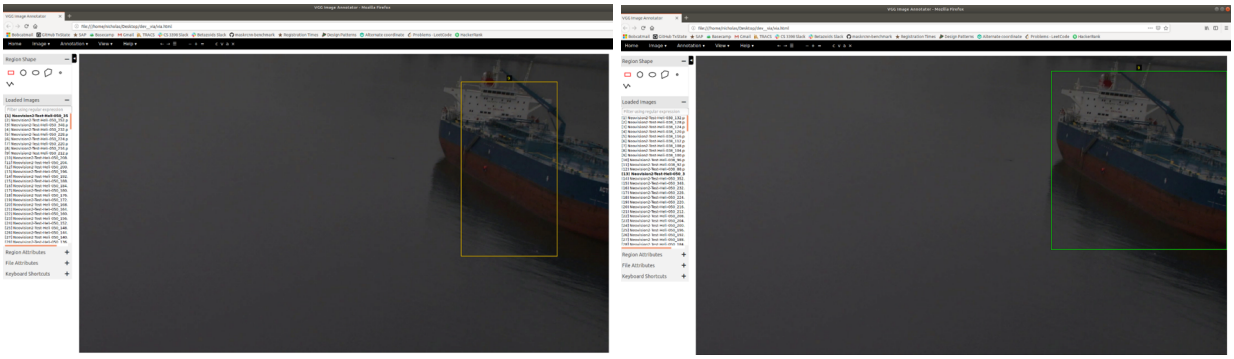

Figure 4. VIA tool adaptation for asset identification and monitoring: Analyst can add new annotation and draw a new bounding box (left). The annotation will automatically label the region that has the with highest IoU score with that bounding box (center). Analyst can choose to refine the label or make it more specific (right).[22]

annotates the asset, as illustrated in Figure 3. This functionality helps us extend the training set for DNN system with all objects whose representation is missing from the training data, and also add new levels of description e.g. boat, target boat. The AIM extended VIA system allows the analyst to add missing annotations, see Figure 3, or refines annotated bounding boxes, see Figure 4. This interactive functionality ensures that the target objects are labeled and marked correctly, and minimizes the effect of edversarial examples on the training and model refining process. Full demo of the tool is available on YouTube .[22] The sample dataset and sample annotations in VIA Annotator example are from DARPA NEOVISION dataset .[26]

4. DEEP NEURAL NETWORK REFINEMENT FOR DOMAIN ADAPTATION

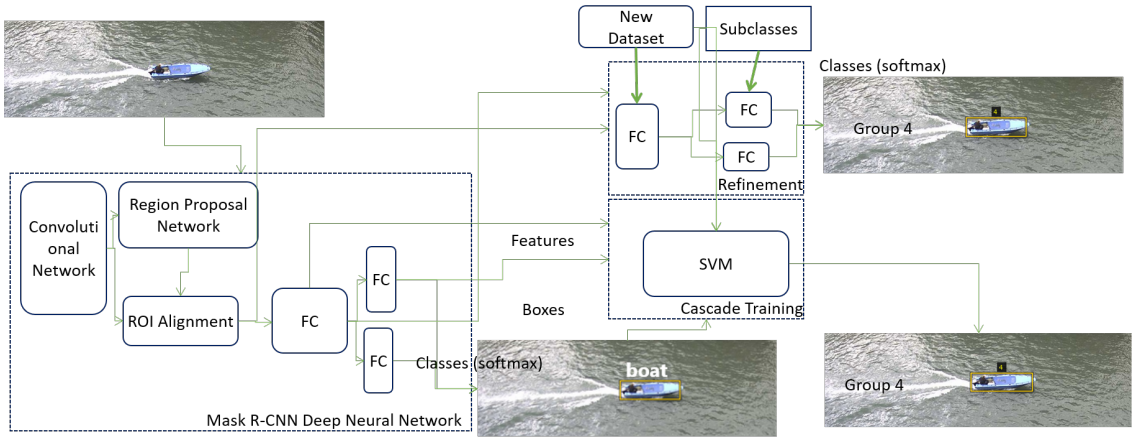

Figure 5. (Top) Deep Neural Network Refinement using small domain specific labels; (Bottom) Cascading classifier learns sub-class models from deep features for regions that have high generic class score.

We utilize the discriminate power of raw features produced by deep neural network system before the classification step, as demonstrated in.[11,12,27] The final form of the features cannot capture aspects that separate one member of a generic class (e.g. car) from another. If an analyst is looking for a specific kind of boat, as illustrated in Figure 4, and labels it as a tugboat, we can use this new specialized labeling to separate what characterizes grey car from all the other cars using underlying DCNN features. In the deep neural network inference phase, We propose two different ways to train the model at multiple level of description:

Deep Neural Network Refinement: we remove the sofmax layer of domain-adapted DNN with refined labels, and train the new network on mission specific labels, as illustrated in Figure 5(top)

NAVAIR Public Release 2019-349. Distribution Statement A – “Approved for public release; distribution is unlimited"

Cascading classifier: we save top region proposal network candidates[4,28,29] and associated high dimensional features for those regions and their generic labels, and we train the multi-classification system to refine the generic label (e.g. boat) while using only deep features for regions that were recognized as boat, as illustrated in Figure 5(bottom).

5. EXPERIMENTS, RESULTS, AND FINDINGS

Collection	COCO		IPATCH		MarDCT	
Dataset	Train	Test	Train	Test	Train	Test
No. frames	118287	5000	3581	7049	2696	846
No. boats	10759	430	6305	8151	3341	846

Table 1. Overview of the publicly available datasets: COCO,[9] IPATCH,[30] and MarDCT[31] data collections.

Datasets use are publicly available consumer and maritime datasets. Number of annotation instances and dataset characteristics are shown in Table 5. Note that IPATCH and MarDCT datasets provided frame or scene-based annotation only. For IPATCH dataset, we utilized annotations from,[19] and for MarDCT we have used our robust boat detector and extended VIA annotator to translate frame-level annotations to object-level annotations, see[22] for examplar pipeline.

COCO benchmark dataset is used as a benchmark for performance evaluation of our transfer learning strategy. COCO, Common Objects in Context dataset consists of images with complex everyday scenes containing common objects in their natural context. COCO dataset contains 91 objects types common in consumer photography, and total of 2.5 million labeled objects in 328k images. We utilize extensive boat annotation in COCO train to imrpove our generic model performance.[9]
IPATCH dataset is our benchmark maritime dataset, and it consists of data collected from multiple sensor surveillance to protect a vessel at sea from piracy. The recordings represent a series of realistic maritime piracy scenarios. In this experiment, we use Low Level Challenge Dataset as IPATCH train, and Mid Level PETS Dataset as out validation set.[32] We utilize IPATCH as part of our multi-source domain adaptation setup.

MarDCT, Maritime Detection, Classification, and Tracking (MarDCT) dataset consists of images coming from multiple sources and from different scenarios. MarDCT classification dataset contains images from 24 different categories of boats navigating in the City of Venice (Italy). We utilize its finer level of annotation to test our system of identifying objects at multiple level of description.[31] The labels, coverage, and how we grouped them for the analysis are shown in Table 3, where exemplars of each of the categories are in Figure 7.

Deep Learning Framework We rely on the baseline pytorch 1.0 implementation of MaskRCNN.[33] Our DNN is created using ResNet50[2] architecture and for each network we train 180,000 epochs. **System** Server with four NVIDIA GeForce GTX 1080 Ti GPUs is used for training and inferencing.

Performance Evaluation is standardized COCO benchmark evaluation metric[9] where True Positive $TP(c)$ for class c as a proposal was made for class c with probability higher than the threshold, and there actually was an object of class c, and the IOU is larger than set threshold. False Positive $FP(c)$ for class c is computed when a proposal was made for class c, but there is no ground truth object of class c. False Negative $FN(c)$ for class c is when a proposal was made for class c, but it is lower than the threshold; or IoU with the ground truth object for class c is lower than than IoU threshold. The average precision (AP) for set IoU is number of true positives over sum of true positives and true negatives, and the recall for the same IoU is number of true positives over sum of true positives and false negatives. Average Precision (Average Recall) is averaging precision (recall) over all classes for specific IoU over a range of IoU thresholds:

$$AP(c) = \frac{|TP(c)|}{(|TP(c)| + |FP(c)|}; Recall(c) = \frac{|TP(c)|}{(|TP(c)| + |FN(c)|}; mAP = \frac{1}{|classes|} \Sigma classes \frac{|TP(c)|}{|TP(c)| + |FP(c)|}$$

NAVAIR Public Release 2019-349. Distribution Statement A – “Approved for public release; distribution is unlimited"

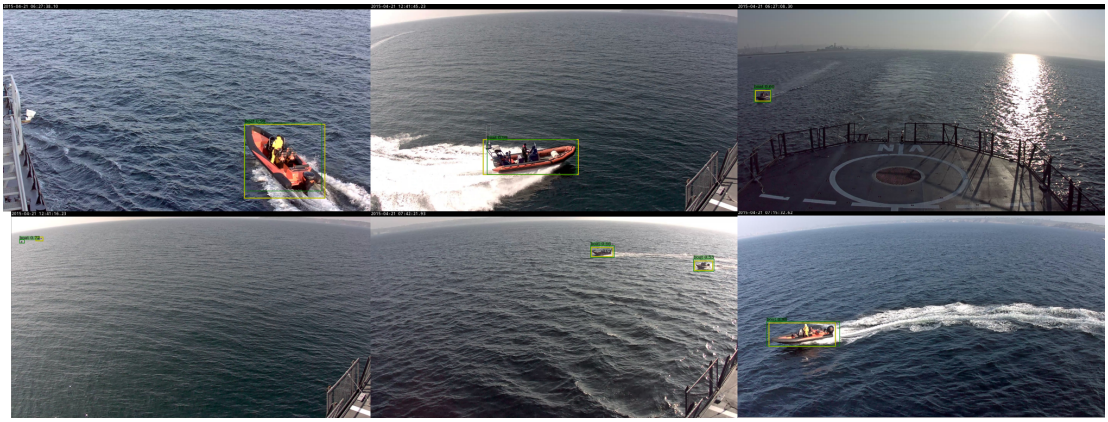

Figure 6. IPATCH Visual evaluation: domain adapted model performs significantly better for small boats and various environments.

Experiment 1: Domain Translation compares the influence of training dataset selection to a model performance. It is an accurate assessment of how features gathered from a large dataset of a different domain could help when applied to a dataset in a more obscure highly variant domain. Performance results are show in Table 2, and few visual exemplars of IPATCH data performance are in Figure 6. Note that model performance significantly improves for IPATCH when domain relevant data is used in training. Even though the training set is much smaller than COCO training set and Validation set, it still offered a significant performance boost. These results show that many of the difficulties in the ocean environment can be captured by creating a dataset that encompasses the domain specific challenges. MarDCT data paints a different picture: all boat objects are centered and relatively large compared to the frame size, and MarDCT data distribution is closer to COCO data distribution: adding domain relevant training examples only marginally improves the precision - we see greater contribution in recall performance, see Table 2 for details.

Table 2. Boat Model Performance measured as average precision on target test set, at threshold set to 0.5

IPATCH Test Set

	Average Precision			Average Recall
Train Set	IoU 0.5	IoU 0.75	IoU 0.5:0.95	IoU 0.5:0.95
COCO Train	0.341	0.106	0.141	0.26
COCO Train + IPATCH Train	**0.925**	0.602	0.747	0.681

MarDCT Test Set

	Average Precision			Average Recall
Train Set	IoU 0.5	IoU 0.75	IoU 0.5:0.95	IoU 0.5:0.95
COCO	0.983	0.979	0.804	0.847
COCO + IPATCH	0.964	0.960	0.788	0.853
COCO + IPATCH + MarDCT	**0.989**	0.989	0.838	0.875

Table 3. MarDCT Boat classes grouped by visual similarity. Note that only 18 out of 24 classes have coverage in test set. Group 1 consists of visually similar "Lanciafino10mBianca", "Lanciafino10m", "Lanciafino10mMarrone", and "Lanciamaggioredi10mBianca" categories, and group 4 encompases the following labels: "Motobarca", "Barchino","Patanella", "Topa", "MotoscafoACTV", "Motopontonerettangolare", "Gondola", "Raccoltarifiuti","Sandoloaremi", "Alilaguna", "Polizia", "Ambulanza"

Collection	Group 1	Group 2	Group 3	Group 4	Other boats	Background
Dataset	Lancia*	VaporettoACTV	Mototopo	See caption	Other boats	Background
Train	598	483	563	730	340	other
Test	296	232	199	241	605	other

NAVAIR Public Release 2019-349. Distribution Statement A – "Approved for public release; distribution is unlimited"

Experiment 2: Refinement and Model Trimming MarDCT dataset is publicly available dataset that provides frame-level labels for 24 types of boats that operate in city of Venice.

The goal of this experiment was to show that a pre-trained model can be used to detect a new class of objects separate from which it was previously trained. In this experiment, we further refined the COCO IPATCH model by adding MarDCT to the training pipeline. After 1000 iterations of refinement the model's final layer was trimmed from the 81 COCO layers (80 + background) to 6 mardct class layers (5 + background), classes are listed in Table 3. The newly trimmed model was then trained on mardct for 100k iterations. The results show that the learned behavior of a boat are still present from before trimming and that the model is able to classify boats according to the four MarDCT classes in the new training as demonstrated in Table 4. The precision of the refinement approach is high for low IoU, and Recall is high accross the board. Figure 7 shows correctly classified examples from MarDCT test data.

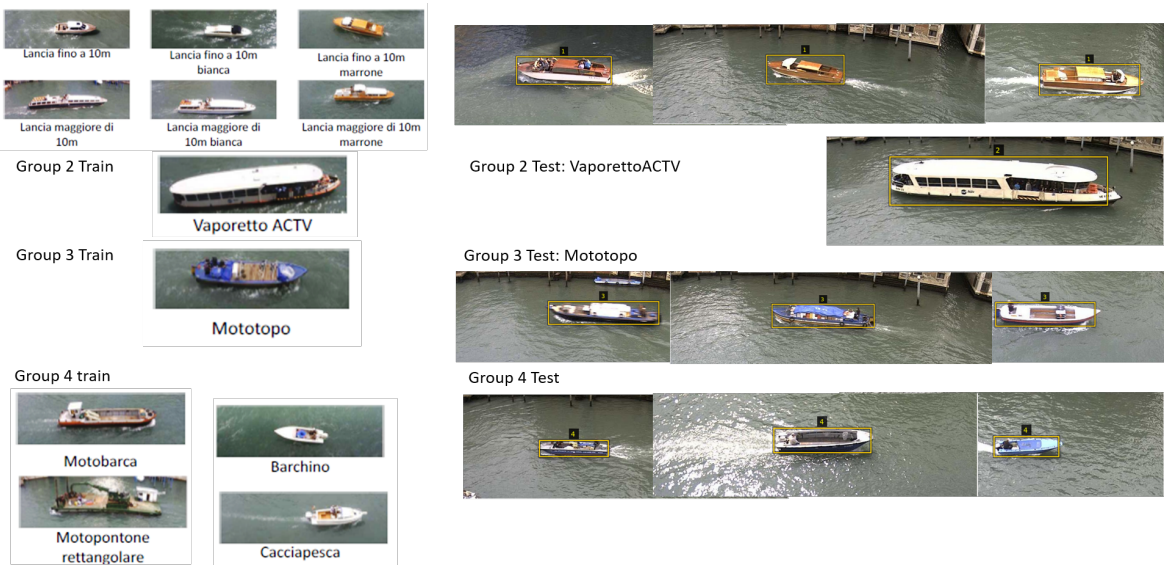

Figure 7. (Left) Training set examples, and (Right) Correctly classified Test examples. Group 1 consists of visually similar "Lanciafino10mBianca", "Lanciafino10m", "Lanciafino10mMarrone", and "Lanciamaggioredi10mBianca" categories, and group 4 encompasses the following labels: "Motobarca", "Barchino","Patanella", "Topa", "MotoscafoACTV", "Motopontonerettangolare", "Gondola", "Raccoltarifiuti","Sandoloaremi", "Alilaguna", "Polizia", "Ambulanza"

Table 4. Four boat sub-classes model refinement performance measured as average precision on target test set, at threshold set to 0.5. We compare them with simple modeling using deep features (* notes that the Deep Features extracted used older version of the Mask RCNN framework.

MarDCT Test Set	Average Precision			Average Recall
COCO + IPATCH network weights	IoU 0.5	IoU 0.75	IoU 0.5:0.95	IoU 0.5:0.95
MarDCT Train refinement	**0.958**	0.903	0.703	**0.766**
SVM*	0.53	N/A	N?A	0.698

Summary We have demonstrated robust way of increasing model performance when adjusted to domain characteristics. The greatest discriminator is domain sensitive training data. Maritime domain lack alternative targets that would be incorrectly associated as maritime vehicles allowed us to relax the parameter constraints learned on urban natural scenes in consumer photos, adjust parameters of the model inference, and achieve robust performance and high precision and recall numbers. We have shown the performance of refinement and cascaded approach for sub-class identification in Table 4. Network refinement seems like a more promising direction.

NAVAIR Public Release 2019-349. Distribution Statement A – "Approved for public release; distribution is unlimited"

6. CONCLUSION

We propose and evaluate an approach for multi-source domain adaptation when few noisy annotations are available. Varying resolution quality of operational data, size of objects of interest, view occlusions, and large variation in sensors due to sheer nature of overhead systems as compared to consumer devices contribute to degradation of the classification and recognition when applied to overhead sensor data. We exploit the domain characteristics to refine the deep learning framework, and show that our transfer learning strategy produces models that reliably and accurately discriminate sea objects from overhead imagery data comparable to consumer data benchmarks. Next, we introduce the notion of modeling at multiple levels of description utilizing deep features and existing deep network weights to learn the difference between sub-categories, and demonstrate superior performance of over 90% precision and 70% recall when enough samples are presented.

ACKNOWLEDGMENTS

This material is based upon work supported by NAVAIR under contracts STTR N68335-16-C-0028 and SBIR N68335-18-C-0199. The views, opinions, and/or findings contained in this article are those of the authors and should not be interpreted as representing the official views or policies of the Department of Defense or the U.S. Government. Authors thank Brent Redmon for Figure 1, and Benjamin Garrard for initial analysis of MarDCT dataset.

REFERENCES

[1] Krizhevsky, A., Sutskever, I., and Hinton, G. E., "Imagenet classification with deep convolutional neural networks," in [*Neural Information Processing Systems (NIPS)*], (2012).

[2] He, K., Zhang, X., Ren, S., and Sun, J., "Deep residual learning for image recognition," in [*The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*], (June 2016).

[3] Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., and Rabinovich, A., "Going deeper with convolutions," *CoRR* (2014).

[4] He, K., Gkioxari, G., Dollár, P., and Girshick, R. B., "Mask r-cnn," *2017 IEEE International Conference on Computer Vision (ICCV)* , 2980–2988 (2017).

[5] Redmon, J., Divvala, S., Girshick, R., and Farhadi, A., "You only look once: Unified, real-time object detection," in [*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*], 779–788 (2016).

[6] Jifeng Dai, Yi Li, K. H. J. S., "R-FCN: Object detection via region-based fully convolutional networks," *arXiv preprint arXiv:1605.06409* (2016).

[7] Everingham, M., Van Gool, L., Williams, C. K. I., Winn, J., and Zisserman, A., "The pascal visual object classes (voc) challenge," *International Journal of Computer Vision* **88**, 303–338 (June 2010).

[8] Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., Berg, A., and Fei-Fei, L., "Imagenet large scale visual recognition challenge," *CoRR* (2014).

[9] Lin, T.-Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., and Zitnick, C. L., [*Microsoft COCO: Common Objects in Context*], 740–755, Springer International Publishing (2014).

[10] Hara, K., Vemulapalli, R., and Chellappa, R., "Designing deep convolutional neural networks for continuous object orientation estimation," *arXiv* (2017).

[11] evikalp, H., Dordinejad, G. G., and Elmas, M., "Feature extraction with convolutional neural networks for aerial image retrieval," in [*2017 25th IEEE Signal Processing and Communications Applications Conference (SIU)*], 1–4 (May 2017).

[12] Sharif Razavian, A., Azizpour, H., Sullivan, J., and Carlsson, S., "Cnn features off-the-shelf: An astounding baseline for recognition," in [*The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*], (June 2014).

[13] Yosinski, J., Clune, J., Bengio, Y., and Lipson, H., "How transferable are features in deep neural networks?," in [*Advances in Neural Information Processing Systems 27*], Ghahramani, Z., Welling, M., Cortes, C., Lawrence, N. D., and Weinberger, K. Q., eds., 3320–3328 (2014).

NAVAIR Public Release 2019-349. Distribution Statement A – "Approved for public release; distribution is unlimited"

[14] Snell, J., Swersky, K., and Zemel, R., "Prototypical networks for few-shot learning," in [*Advances in Neural Information Processing Systems 30*], Guyon, I., Luxburg, U. V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., and Garnett, R., eds., 4077–4087 (2017).

[15] Maggiori, E., Tarabalka, Y., Charpiat, G., and Alliez, P., "Convolutional neural networks for large-scale remote-sensing image classification," *IEEE Transactions on Geoscience and Remote Sensing* **55**, 645–657 (Feb 2017).

[16] Sumbul, G., Cinbis, R. G., and Aksoy, S., "Fine-grained object recognition and zero-shot learning in remote sensing imagery," *IEEE Transactions on Geoscience and Remote Sensing* **56**, 770–779 (Feb 2018).

[17] Bosch, M., Christie, G., and Gifford, C., "Sensor adaptation for improved semantic segmentation of overhead imagery," in [*2019 IEEE Winter Conference on Applications of Computer Vision (WACV)*], 648–656 (Jan 2019).

[18] Alexe, B., Deselaers, T., and Ferrari, V., "Measuring the objectness of image windows," *IEEE Transactions of Pattern Analysis and Machine Intelligence (TPAMI)* (2012).

[19] Warren, N., Garrard, B., Staudt, E., and Tesic, J., "Transfer learning of deep neural networks for visual collaborative maritime asset identification," in [*2018 IEEE 4th International Conference on Collaboration and Internet Computing (CIC)*], 246–255 (Oct 2018).

[20] Naphade, M., Smith, J. R., Tesic, J., , Hsu, W., Kennedy, L., Hauptmann, A., and Curtis, J., "Large-scale concept ontology for multimedia," *IEEE MultiMedia* **13**, 86–91 (July 2006).

[21] Hoffman, J., Guadarrama, S., Tzeng, E., Hu, R., Donahue, J., Girshick, R., Darrell, T., and Saenko, K., "LSDA: Large scale detection through adaptation," in [*Neural Information Processing Systems (NIPS)*], (2014).

[22] Warren, N. and Tesić, J., "Aim extension for via annotator." https://youtu.be/b22mGCgcTQY (2019).

[23] Akhtar, N. and Mian, A., "Threat of adversarial attacks on deep learning in computer vision: A survey," *IEEE Access* **6**, 14410–14430 (2018).

[24] Swirszcz, G., Czarnecki, W. M., and Pascanu, R., "Local minima in training of neural networks," *arXiv e-prints* (Nov 2016).

[25] Dutta, A., Gupta, A., and Zissermann, A., "VGG image annotator (VIA)." http://www.robots.ox.ac.uk/ vgg/software/via/ (2016). Version: 2.0.6, Accessed: Mar 25 2019.

[26] Kasturi, R., Goldgof, D. B., Ekambaram, R., Pratt, G., Krotkov, E., Hackett, D. D., Ran, Y., Zheng, Q., Sharma, R., Anderson, M., Peot, M., Aguilar, M., Khosla, D., Chen, Y., Kim, K., Elazary, L., Voorhies, R. C., Parks, D. F., and Itti, L., "Performance evaluation of neuromorphic-vision object recognition algorithms," in [*2014 22nd International Conference on Pattern Recognition*], 2401–2406 (Aug 2014).

[27] Huang, F. J. and LeCun, Y., "Large-scale learning with svm and convolutional for generic object categorization," in [*2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*], **1**, 284–291 (June 2006).

[28] Agrawal, P., Girshick, R., and Malik, J., "Analyzing the performance of multilayer neural networks for object recognition," in [*Proceedings of the European Conference on Computer Vision (ECCV)*], (2014).

[29] Ren, S., He, K., Girshick, R., and Sun, J., "Faster r-cnn: Towards real-time object detection with region proposal networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence* **39**, 1137–1149 (June 2017).

[30] Patino, L., Nawaz, T., Cane, T., and Ferryman, J., "Pets 2017: Dataset and challenge," in [*2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*], 2126–2132 (July 2017).

[31] Bloisi, D. D., Iocchi, L., Pennisi, A., and Tombolini, L., "ARGOS-Venice boat classification," in [*Advanced Video and Signal Based Surveillance (AVSS), 2015 12th IEEE International Conference on*], 1–6 (2015).

[32] Patino, L., Cane, T., Vallee, A., and Ferryman, J., "Pets 2016: Dataset and challenge," in [*The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*], (June 2016).

[33] Massa, F. and Girshick, R., "maskrcnn-benchmark: Fast, modular reference implementation of Instance Segmentation and Object Detection algorithms in PyTorch." https://github.com/facebookresearch/maskrcnn-benchmark (2018). Accessed: March 25 2019.

NAVAIR Public Release 2019-349. Distribution Statement A – "Approved for public release; distribution is unlimited"