



ARL-TR-8508 • SEP 2018



Developing Neural Scene Understanding for Autonomous Robotic Missions in Realistic Environments

by Arnold D Tunick and Ronald E Meyers

Approved for public release; distribution is unlimited.

NOTICES

Disclaimers

The findings in this report are not to be construed as an official Department of the Army position unless so designated by other authorized documents.

Citation of manufacturer's or trade names does not constitute an official endorsement or approval of the use thereof.

Destroy this report when it is no longer needed. Do not return it to the originator.



Developing Neural Scene Understanding for Autonomous Robotic Missions in Realistic Environments

by Arnold D Tunick and Ronald E Meyers
Computational and Information Sciences Directorate, ARL

REPORT DOCUMENTATION PAGE			Form Approved OMB No. 0704-0188		
Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing the burden, to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.					
1. REPORT DATE (DD-MM-YYYY) September 2018		2. REPORT TYPE Technical Report		3. DATES COVERED (From - To) May 2016–Aug 2018	
4. TITLE AND SUBTITLE Developing Neural Scene Understanding for Autonomous Robotic Missions in Realistic Environments			5a. CONTRACT NUMBER		
			5b. GRANT NUMBER		
			5c. PROGRAM ELEMENT NUMBER		
6. AUTHOR(S) Arnold D Tunick and Ronald E Meyers			5d. PROJECT NUMBER		
			5e. TASK NUMBER		
			5f. WORK UNIT NUMBER		
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) US Army Research Laboratory Computational and Information Sciences Directorate (ATTN: RDRL-CII-A) 2800 Powder Mill Road Adelphi, MD 20783-1138			8. PERFORMING ORGANIZATION REPORT NUMBER ARL-TR-8508		
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)			10. SPONSOR/MONITOR'S ACRONYM(S)		
			11. SPONSOR/MONITOR'S REPORT NUMBER(S)		
12. DISTRIBUTION/AVAILABILITY STATEMENT Approved for public release; distribution is unlimited.					
13. SUPPLEMENTARY NOTES					
14. ABSTRACT Achieving rapid and robust understanding of scenes in complex and changing environments is critically important for autonomous robotic systems performing tasks to support realistic outdoor missions. Outdoor missions are scenario-oriented and are not just dependent on searches for objects. Autonomous missions need to detect time-sensitive scenarios that are signified by salient objects and environments providing mutual context in an evolving scene. Real-world, time-sensitive mission scenarios may include human activity and evolving threats, such as forest fires, chemical hazards, or adversary encroachment. We describe the implementation of a novel composite neural software configuration for complex outdoor missions that is tested against simulated test mission scenarios using two convolutional neural network (CNN) models separately trained on component objects and places image databases. Our proof-of-principle testing of the composite CNN supports the benefits of such a system when tuned to detecting time-sensitive scenarios that are keyed to the success of the mission. We examine five real-world mission scenarios to analyze the benefits of adding environmental data assimilation and physical modeling to improve neural scene understanding. We find that achieving autonomous inference of mission intent from detected and surveilled activities would make autonomous missions even more valuable.					
15. SUBJECT TERMS deep learning, scene exploration, dynamic environment, object detection, place detection, convolutional neural networks					
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT UU	18. NUMBER OF PAGES 47	19a. NAME OF RESPONSIBLE PERSON Arnold D Tunick
a. REPORT Unclassified	b. ABSTRACT Unclassified	c. THIS PAGE Unclassified			19b. TELEPHONE NUMBER (Include area code) (301) 394-1233

Contents

List of Figures	v
List of Tables	vi
Acknowledgments	vii
1. Introduction: Need for Autonomous Neural Networks for Outdoor Missions	1
1.1 Neural Scene Understanding	1
1.2 Military and Time-Sensitive Outdoor Missions	1
1.3 Mitigating Uncertainty Introduced by Changing Weather, Terrain, and Environmental Surroundings	2
1.4 Outdoor Mission Scenarios	3
2. Related Neural Network Model Work	3
3. Neural Network Models, Training, and Concepts	4
3.1 Object-Trained Neural Network Model	5
3.2 Places-Trained Neural Network Model	7
4. Autonomous Detection of Test Mission Scenarios	8
4.1 Cost–Benefit Considerations in Operational Implementation of Neural Network Search Strategies	15
4.2 Computer Processing, Timing, and Automation	16
5. Bringing About Rapid and Robust Autonomous Scene Exploration	17
5.1 Dynamic Environmental Data Retrieval	17
5.2 Physics-Based Modeling	19
5.3 Integrated Scene Understanding	20
5.3.1 Ground Search and Rescue	20
5.3.2 Aerial Reconnaissance	20
5.3.3 Nuclear, Biological, and Chemical (NBC) Hazard Detection	21

5.3.4	Cave and Tunnel Reconnaissance	21
5.3.5	Sniper Detection	21
5.4	Applicability to Diverse Evolving and Time-Sensitive Missions	22
6.	Rapid and Robust Neural Detection Scheme Models	24
6.1	Autonomous Inference of Mission Intent	27
7.	Conclusions	28
8.	References	29
	List of Symbols, Abbreviations, and Acronyms	37
	Distribution List	38

List of Figures

Fig. 1	Representative object-trained CNN model result that illustrates the need to include the recognition of dynamic environment image clues and context in the scene understanding approach (e.g., weather conditions and trends, visibility, and terrain). The test image is annotated with the top-5 most likely semantic labels and corresponding top-5 confidence levels for the neural net prediction. ...	5
Fig. 2	Representative results from the Theano-AlexNet CNN software implementation using test images that depict various dynamic environmental features. Each test image is annotated with the top-5 most likely semantic labels and corresponding top-5 confidence levels.	6
Fig. 3	Example of Places-CNN model output for selected test images shown in Fig. 2. Each test image is annotated with the top-5 most likely semantic labels and corresponding top-5 confidence levels as well as a list of scene attributes (as discussed in the text).	7
Fig. 4	Representative composite CNN results for the simulated test mission scenario “finding Army tanks in specific environmental settings”. Top-5 image classification labels and top-5 probabilities are annotated for the object-centric trained CNN (above) and location/places-centric trained CNN (below).	11
Fig. 5	Representative composite CNN results for the autonomous detection of the simulated test mission scenario “finding persons in specific environmental settings”. Top-5 image classification labels and top-5 probabilities are annotated for the object-trained CNN (above) and location/places-trained CNN (below). (Top) find a person on the beach; on a boardwalk; on a bridge, in a creek; on a crosswalk. (Middle): find a person in the desert; in a wheat field; on a forest path; on a golf course; on the highway. (Bottom): find a person on a mountain path; on a railroad track; on a ski slope; in a snow field; in a trench.	13
Fig. 6	Composite CNN model results for the autonomous detection of the simulated test mission to find a baseball player playing baseball (i.e., a baseball game scenario). Top-5 image classification labels and top-5 probabilities are annotated for the object-centric trained CNN (above) and location/places-centric trained CNN (below).	14
Fig. 7	Example of the composite CNN model output where there is a false positive occurrence in the detection of a simulated test mission scenario using the extended selection criteria described in the text....	15
Fig. 8	Robotic platform with mounted camera at ARL.....	24
Fig. 9	Robotic neural detection of a person behind tree branches. The test image was obtained from video recorded on a robotics platform at ARL.	25

Fig. 10	Panoramic robotic neural detection of a person behind tree branches. Note the robot track in the near-field. The test image was obtained from video recorded on a robotics platform at ARL.....	25
Fig. 11	Neural detection of persons (Soldiers) in camouflage. The test image was obtained from the Internet.	25
Fig. 12	Neural detection of a person (Soldier) in camouflage and vegetative terrain. The test image was obtained from the Internet.....	26
Fig. 13	Neural detection of a person (Soldier) in vegetative terrain. The test image was obtained from the Internet.	26
Fig. 14	Neural detection of persons (Soldiers) in obscurants. The test image was obtained from the Internet.	27

List of Tables

Table 1	Summary of representative scenarios for simulated test missions and real-world, time-sensitive missions.....	9
Table 2	Composite CNN model results for the autonomous detection of Army tanks in specific environmental settings.....	10
Table 3	Composite CNN model results for the autonomous detection of persons in specific environmental settings	12
Table 4	Composite CNN model results for the autonomous detection of a baseball player playing baseball	14
Table 5	Summary of the composite CNN model results for the autonomous detection of simulated test mission scenarios	15
Table 6	Integrated scene understanding approach for five representative real-world, time-sensitive missions (summary)	23

Acknowledgments

This research was supported by the US Army Research Laboratory.

1. Introduction: Need for Autonomous Neural Networks for Outdoor Missions

1.1 Neural Scene Understanding

Neuron cells in the brain enable us to understand scenes by assessing the spatial, temporal, and feature relations of objects in the scene to the environment and ourselves. Clarifying how humans gain scene understanding is an ongoing active area of theoretical and experimental research (Biederman 1987; Bar 2004; Battaglia et al. 2013; Sofer et al. 2015; Malcolm et al. 2016) that even extends to experiments that relate magnetic resonance imaging and other sensor measurements of brain neural activity to images of scenes viewed by subjects (Aminoff and Tarr 2015; Aminoff et al. 2015; Suave et al. 2017). To gain a better understanding of neural information processing, experiments and analyses related to imaging biological neural networks and electrical neural activity in the brain also have been reported (Hall et al. 2012; Karadas et al. 2017). Despite the complexity of mimicking human neural activity and because of the potential for enormous gains, there has been intense effort to use computer neural networks to augment human neural intelligence to improve our scene understanding (Krizhevsky et al. 2012; Zhou et al. 2014, 2017; Karpathy and Li 2015; Lim et al. 2017; Qiao et al. 2017; Yu et al. 2017; Yang et al. 2017; Han et al. 2017). In the age of robots and drones, achieving rapid and robust understanding of scenes has become a critical goal for autonomous robotic systems performing tasks to support realistic outdoor missions in complex and changing environments (ARL 2014; Judson 2016; Sustersic et al. 2016). Considering all types of environments and terrain is an essential requirement that has been included in mission planning strategies for a long time (AWC 2001). In this study, we also examine the potential for achieving autonomous inference of mission intent from detected and surveilled activities, which would make autonomous missions even more valuable.

1.2 Military and Time-Sensitive Outdoor Missions

Outdoor missions tend to be complex and time sensitive, and involve changing circumstances, such as those related to search and rescue, safety, firefighting, and air-pollution hazards. Military operations are exemplars of complex outdoor missions, in particular, those where there is a need to reduce the analytical burden to process time-sensitive information on an information-saturated and rapidly changing battlefield (Howard and Cambria 2013). Lessons learned in our following analysis of five representative real-world, time-sensitive missions can be applied to countless outdoor domestic and defense endeavors.

1.3 Mitigating Uncertainty Introduced by Changing Weather, Terrain, and Environmental Surroundings

Scene understanding for realistic outdoor missions remains an unsolved problem (RCTA 2012; Piekiewicz et al. 2016; Tai and Liu 2017) due to the uncertainty of inferring the mutual context of detected objects and the changing weather, terrain, and environmental surroundings. While gaining scene understanding is challenging even in static surroundings, dynamic applications such as terrain and obstacle traversal in changing environments greatly complicates the missions. The application of neural networks to realistic missions requires attention to changing weather and terrain features encountered along paths that can introduce uncertainty in neural predictions. For detecting objects in a scene, neural network models can be used to predict the most likely semantic labels or categories of objects by a calculated probability or confidence score (Krizhevsky et al. 2012; Zhou et al. 2014, 2017). By predicting the activity of neural systems in humans for sensory processing or behavioral control, neural predictions also can mean the output from neuroscience models (Gallant and David 2015). In this report, the term “neural prediction” refers to a neural network model solution at a given time rather than a forecast of what may occur at a future time, as with physics-based models.

Neural predictions can be combined with environmental data assimilation (Tunick 2016; Yang 2017) and physics-based models (Battaglia et al. 2013; Ullman et al. 2014; Fragkiadaki et al. 2016; Wu et al. 2015; Lake et al. 2016; Zhang et al. 2016) to enable improved scene understanding. By improved scene understanding, we mean the additional information from recorded images that can be gained by the integration of neural network software, multiple training data sets, environmental data assimilation, and physical modeling for robotic systems performing tasks to support realistic outdoor missions. As an example, an automated vision system may readily detect changes in the ground surface along its path as a new or different object in the field of view. It would be a challenge for neural networks alone to be able to differentiate such features as shallow or deep water; thick, thin, or melting ice; freezing rain; blowing dust and sand; or snow, mud, or quicksand if they had not been encountered in the training data sets. Our instructive analysis of five representative real-world, time-sensitive missions shows that adding physics-based modeling and dynamic environmental data such as terrain, morphology, weather, visibility, and illumination to neural network training can minimize unpredictability and thereby constrain neural predictions to physically realizable solutions.

1.4 Outdoor Mission Scenarios

Outdoor missions are scenario-oriented and are not just dependent on searches for objects, as is the goal of most scene understanding applications. Scenarios are signified by salient objects and environments providing mutual context in an evolving scene. Existing scene understanding applications, while individually limited, do provide many useful components for improved outdoor scene exploration. To be most useful for autonomous outdoor missions, robotic scene understanding systems need to detect time-sensitive and evolving scenarios, not just objects. Real-world, time-sensitive mission scenarios may include human activity and evolving threats, such as forest fires, chemical hazards, or adversary encroachment. In this report, the implementation of a novel composite neural software configuration for complex outdoor missions is described and tested against simulated test mission scenarios using two convolutional neural network (CNN) models separately trained on component objects and places image databases. A simulated test mission scenario is a scenario whose detection is currently enabled using a composite CNN trained on available object and places image databases. The real-world, time-sensitive mission scenarios discussed here are any scenarios or activities that are keyed to the success of the outdoor mission. Robotic systems supporting real-world autonomous outdoor missions tend to look for scenarios comprising activities or capabilities. A scenario activity might involve a robotic system looking for a sniper who has a gun and is trying to surveil potential targets. A scenario capability might mean a robotic system assessing a chemical hazard event that occurs near a population center, where there are encroaching fire and weather activities that may distribute the toxins into the population.

2. Related Neural Network Model Work

Previously, Xiang et al. (2002) and Gori et al. (2016) reported on experiments conducted using indoor video images to autonomously search for and identify human activities that happen in sequence or concurrently. However, the methods of Xiang et al. (2002) and Gori et al. (2016) did not address neural network models or outdoor environments, as we do. Li and Li (2007) discussed the classification of human activities or events in static images using both object and scene categorizations; however, Li and Li (2007) implemented detection models based on latent Dirichlet allocation (Blei et al. 2003) rather than neural networks. Daniels and Metaxas (2018) also discussed scenarios as a way of representing scenes using neural networks. The methods of Daniels and Metaxas (2018) predict the most likely scenarios in a given setting that result from the combined detection of multiple objects in the scene. While the methods of Daniels and Metaxas (2018) demonstrate practical scene exploration, they do not address the detection of

evolving scenarios to support time-sensitive outdoor missions, as described in this report.

Wang et al. (2015, 2017) reported on event recognition using a combination of object and scene neural networks to recognize cultural events or activities depicted in still images, such as carnivals, marathons, parades, and holiday celebrations (Baro et al. 2015). However, the methods of Wang et al. (2015, 2017) did not address a more comprehensive structure for scene understanding that can be achieved by integrating environmental data, physics-based modeling, and neural networks to improve the next generation of scene understanding software tools. As an example, Yang (2017) reported on the integration of recorded sensor data, physical modeling, and neural network analysis to investigate chemical dispersion and chemical hazard assessment.

3. Neural Network Models, Training, and Concepts

In this section, neural network models, training, and concepts are reviewed. Much research has focused on the concept of saliency, as if the importance of an object to us or our missions depends solely on how much it visually stands out in a scene. To be useful for outdoor missions, this is not sufficient since the environment and mission objectives must be considered. The concept of saliency estimation has been helpful to computationally identify elements in a scene that immediately capture the visual attention of an observer (Itti et al. 1998; Xu et al. 2010; Perazzi et al. 2012). Several recent papers have discussed concepts associated with visual saliency to enhance automated navigation and scene exploration (Roberts et al. 2012; Yeomans et al. 2015; Warnell et al. 2016). However, the most active or salient object(s) in a scene, by this definition, may not represent the most important or meaningful feature(s) of the scene. Environmental factors such as changing illumination, precipitation, and vegetation can modify saliency and the context of an outdoor scene, obscure features, and significantly degrade object recognition (Wohler and Anlauf 2001; Narasimhan and Nayar 2002; Pepperell et al. 2014; Sunderhauf et al. 2015; Neubert and Protzel 2016; Valada et al. 2016). When applied to neural network object or place recognition, an autonomous robotic system may predict the correct class label for the principal object shown in a test image (Krizhevsky et al. 2012; Zhou et al. 2014, 2017; Karpathy and Li 2015), yet overlook key environmental features that can provide important information related to the outdoor mission. Figure 1 shows that an object-trained neural net model correctly classified an image of a tank (in the near-field view) with high confidence. In the far-field of the image, a tan background was indicative of a sand storm. While the neural network training in this case produced a single label for object recognition, it was not sufficient to identify a potentially vital piece of

environmental information that otherwise may have been recognized by a human observer. The consequence of low visibility can impede many types of navigation, reconnaissance, and target acquisition, and blowing dust or sand can decrease the effectiveness of embedded equipment and personnel.



Fig. 1 Representative object-trained CNN model result that illustrates the need to include the recognition of dynamic environment image clues and context in the scene understanding approach (e.g., weather conditions and trends, visibility, and terrain). The test image is annotated with the top-5 most likely semantic labels and corresponding top-5 confidence levels for the neural net prediction.

The inability of the neural network model to identify the sand feature in the previous example is likely related to the problem of domain transfer (You et al. 2015; Zhang et al. 2016). The domain transfer problem is where knowledge learned by neural networks trained on a data set containing images of a particular domain cannot be transferred easily to produce an outcome that is outside the training data set. Neural network models trained on multiple and diverse data sets can mitigate the domain transfer problem and improve the application of scene understanding software for use in dynamic environments.

3.1 Object-Trained Neural Network Model

In this section, the implementation of the Theano-AlexNet CNN model (Krizhevsky et al. 2012; Ding et al. 2015) is presented. The Theano-AlexNet CNN was trained to detect 1,000 different object categories or classes using the ImageNet Large Scale Visual Recognition Challenge 2012 (ILSVRC2012) data set (Russakovsky et al. 2015). All of the 1.2 million training images (i.e., 5004 mini-batches of 256 images) were processed during each of 65 training cycles. The training of the Theano-AlexNet CNN software achieved 78.7%–79.7% accuracy for the top-5 class labels. These training accuracies are in close agreement with those reported by Krizhevsky et al. (2012) and Ding et al. (2015) (i.e., a top-5 accuracy of 81.8% and 80.1%, respectively). For the evaluation of neural network

models, the top-5 accuracy rate is defined as the fraction of test images for which the correct class label is among the five most likely class labels determined by the model (Krizhevsky et al. 2012; Zhou et al., 2014, 2017).

In Fig. 2, representative results from the Theano-AlexNet CNN software implementation are presented using test images that depict various dynamic environmental features. Addressing dynamic environments in scene understanding software can be useful for autonomous systems operating in realistic outdoor environments. As an example, adverse weather may darken images, making it difficult for conventional neural networks to interpret the outdoor scene. However, proper characterization of weather and environmental features using multiple neural network models and training can enhance autonomous scene understanding for the support of outdoor missions.

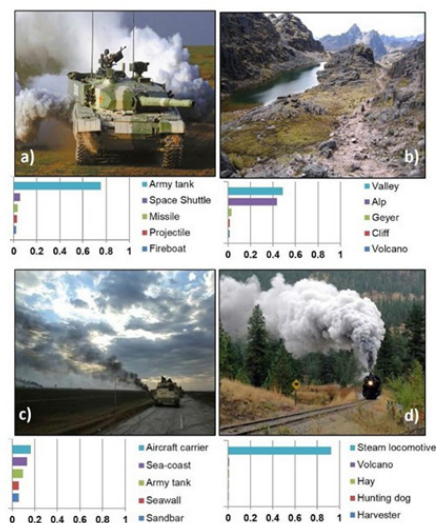


Fig. 2 Representative results from the Theano-AlexNet CNN software implementation using test images that depict various dynamic environmental features. Each test image is annotated with the top-5 most likely semantic labels and corresponding top-5 confidence levels.

In Fig. 2, the neural network model outputs the top-5 most likely class labels and corresponding confidence levels (i.e., top-5 probabilities) for four test images taken from the ILSVRC2012 data set. To do this, the model code was modified and incorporated an inference calculation to extract the desired output (i.e., the $p_{y_given_x}$ output from the CNN softmax layer [Ma 2016]). Figure 2 shows that the neural network model predicted the correct class/semantic labels for the principal object(s) shown in the test images, generally with high confidence. However, a person viewing these images would likely see several additional features, such as clouds, haze, smoke plumes/exhaust, sandy soil, rocky terrain, mountains, river water, trees, and forests. Identifying these environmental features

can be essential to scene understanding, especially when the images are viewed in the context of mission impact. Obscured visibility due to smoke plumes, low illumination due to cloud cover, or difficult navigation due to rocky or forested terrain are all potentially vital pieces of information that can provide enhanced operational awareness. As an example, the tank exhaust and trailing smoke cloud in Fig. 2a are indicative of objects moving in an active scene. Similarly, Figs. 2c and Fig. 2d depict an Army tank on a roadway and a locomotive on tracks, respectively, each with a trailing smoke plume. However, the Army tank in Fig. 2c was identified with a much lower probability (i.e., 9.4%) than the locomotive in Fig. 2d (i.e., 92.4%), even though both objects appear set back from the near-field view. Low illumination of the scene and low visibility of the object in Fig. 2c likely affected the CNN model solution.

3.2 Places-Trained Neural Network Model

In this section, a comparative test image analysis is presented using a places-trained CNN model (Zhou et al. 2014, 2017a), which is similarly based on Krizhevsky et al. (2012), but is trained instead on a places/locations image database. An updated version of the Places-CNN online scene recognition demo was used (Zhao et al. 2017b) to predict the most likely scene categories and scene attributes for several test images depicting various outdoor settings and environmental features. The Places-CNN scene categories are currently based on a 1.8 million image subset of the Places 10 million image data set and the list of scene attributes are based on the data set discussed by Xiao et al. (2010) and Patterson and Hays (2012). Figure 3 provides an example of Places-CNN model output using selected images shown in Fig. 2.

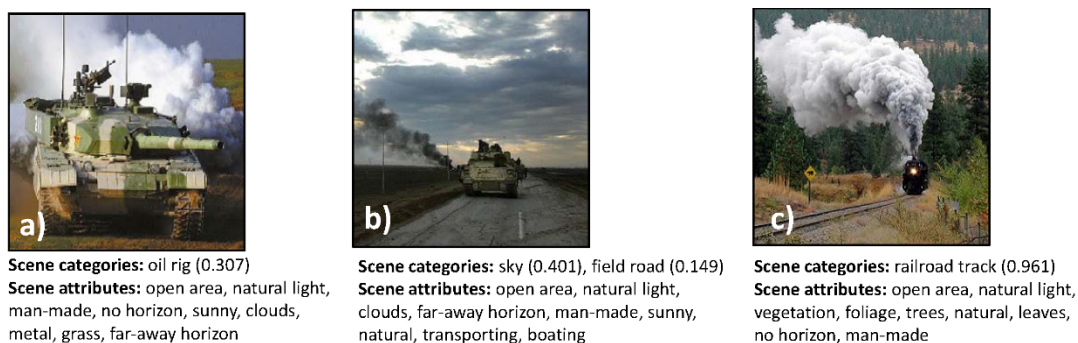


Fig. 3 Example of Places-CNN model output for selected test images shown in Fig. 2. Each test image is annotated with the top-5 most likely semantic labels and corresponding top-5 confidence levels as well as a list of scene attributes (as discussed in the text).

The class label predications from the object- and places-trained neural network models show that dynamic environmental features can impact how a neural network models interpret outdoor scenes.

4. Autonomous Detection of Test Mission Scenarios

Consider the following simulated test mission scenario: An operations center has neural network models trained on million image data sets for both objects as well as environmental settings (i.e., places and/or locations). Skilled personnel use these neural scene understanding software tools to autonomously sift through a large volume of images obtained via remote sensing from robots or drones in order to identify potential threats to personnel or equipment. The neural net models trained on both images of objects and places are configured to search through the recorded image data set and identify all of the images that satisfy predefined criteria for the detection of a mission designated activity or scenario (e.g., “find a tank in the forest” or “find a tank in the desert”). If the object and place recognition criteria are defined so as to increase the number of detected scenarios while minimizing the occurrence of false positives, then the operations center will have used its analyst’s time and resources effectively, enabling a rapid response and dispatch of its mission assets to the targeted scene.

For this section, proof-of-principle experiments were conducted to demonstrate the recognition of three simulated test mission scenarios by identifying both salient objects in addition to relevant environmental settings characteristic of the scenario. To do this, the object-trained CNN was used together with the places-trained CNN¹ to autonomously detect the simulated test mission scenarios shown in Table 1. The analysis of CNN model results for scenario detection proceeded as follows. First, a scenario was considered to be “detected” only when the object-trained and places-trained neural networks jointly determined the top-1 semantic label predictions for both the primary object of an outdoor scene and its environmental setting. The top-1 label prediction is defined as the most likely semantic label determined by the neural net model (Krizhevsky et al. 2012). Through further experiments, the tradeoff between detection constraints and the number of successful scenario detections and number of false positives was considered. For this study, the selection criteria for detected scenarios were expanded to include the top-1 through top-5 predictions, respectively. Table 2 presents a summary of the top-1 through top-5 predictions in separate tallies for the number of objects recognized, the number of places recognized, the number of scenarios detected, and the number of

¹ For the CNN model output shown in Figs. 4, 6, and 7, an earlier version of the Places-CNN online demo was used (<http://places.csail.mit.edu/demo.html>).

false positives. A false positive, for example, would be if the neural net predicted a tank (object) and a desert (place), but the image depicted neither a tank nor a desert. A total of 78 images of Army tanks in different outdoor settings were analyzed and 13–32 scenarios of interest were detected with no false positives, even though in general false positives could occur. Figure 4 shows representative composite CNN results for the scenario “finding a tank in specific environmental settings”. In Fig. 4, the object- and places-CNN model data shown above and below the images for the “tank in the desert”, “tank on the beach”, “tank in a field”, and “tank in a city” cases provide examples of scenarios detected using the top-1 prediction criteria. The other three cases illustrate how an expanded selection criteria can be applied to scenario detections (e.g., “Army tank” was in the top-2 most likely predictions for the “tank in the snow” and “tank in the forest” examples).

Table 1 Summary of representative scenarios for simulated test missions and real-world, time-sensitive missions

Missions		Scenarios					
Simulated test missions	Find a tank in specific environmental settings	Tank in a desert	Tank on a beach	Tank in the snow	Tank in the forest	Tank in a grass field	Tank in the city
	Find a person in specific environmental settings	Person on a beach	Person on a boardwalk	Person on a bridge	Person in a creek	Person in a crosswalk	Person in a desert
		Person on a forest path	Person in a wheat field	Person on a golf course	Person on a highway	Person on a mountain path	Person on a playground
		Person on a railroad track	Person on a ski slope	Person in a snow field	Person in a trench	...	...
Real-world, time-sensitive missions	Find a baseball player playing baseball	Baseball player on a baseball field	...	...	...	...	...
	Ground search and rescue	Find a person in a mission designated weather and terrain setting.					
	Aerial reconnaissance	Find persons or structures in mission designated weather and terrain setting.					
	NBC hazard detection	Identify and quantify NBC hazards in a mission designated area.					
	Cave and tunnel reconnaissance	Identify traversable pathways through an underground mission designated area.					
	Sniper detection	Locate a person with a gun in a window or on a rooftop in a mission designated weather and terrain setting.					

Table 2 Composite CNN model results for the autonomous detection of Army tanks in specific environmental settings

Simulated test mission scenario	No. of objects recognized					No. of places recognized					No. of scenarios detected					No. of false positives
	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5	1–5
Top-1 through Top-5 criteria																
1. tank in the desert (24)	16	17	17	17	17	12	15	16	18	18	8	9	10	12	12	0
2. tank on a beach (10)	6	6	6	6	6	0	5	6	9	10	0	2	3	6	6	0
3. tank in the snow (10)	2	3	4	5	6	1	4	6	6	6	0	1	1	2	2	0
4. tank in the forest (12)	2	3	3	4	5	3	3	4	6	6	0	1	1	2	3	0
5. tank in a field (12)	11	11	11	11	12	5	8	8	8	8	4	7	7	7	8	0
6. tank in the city (10)	6	6	6	6	7	1	1	1	1	1	1	1	1	1	1	0
TOTAL (78)	43	46	47	49	53	22	36	41	48	49	13	21	23	30	32	0

	TANKS IN THE DESERT		TANKS ON THE BEACH		TANKS IN THE SNOW		TANKS IN THE FOREST		TANKS IN A GRASS FIELD		TANKS IN THE CITY	
OBJECT CNN	army tank	0.88	army tank	0.91	amphibious vehicle	0.37	half track	0.33	army tank	0.97	army tank	0.21
	camel	0.04	amphibious vehicle	0.03	army tank	0.36	army tank	0.25	half track	0.01	half track	0.13
	ibex	0.04	plow	0.02	submarine	0.10	amphibious vehicle	0.09	harvester	0.01	palace	0.07
	ship wreck	0.03	oxcart	0.01	cannon	0.02	reel	0.05	amphibious vehicle	0.01	monumental archway	0.06
	gazelle	0.02	---	---	half track	0.02	rifle	0.04	---	---	horse cart	0.05
PLACES CNN												
	desert sand	0.71	ocean	0.33	snowy mountain	0.11	rain forest	0.11	pasture	0.51	medina	0.12
	excavation	0.06	coast	0.24	ski resort	0.08	fountain	0.11	wild field	0.11	slum	0.09
	---	---	sand bar	0.07	boardwalk	0.07	swamp	0.11	wind farm	0.07	castle	0.07
	---	---	harbor	0.06	valley	0.05	trench	0.05	runway	0.06	plaza	0.07
	---	---	boat deck	0.05	---	---	---	---	---	---	palace	0.07

Fig. 4 Representative composite CNN results for the simulated test mission scenario “finding Army tanks in specific environmental settings”. Top-5 image classification labels and top-5 probabilities are annotated for the object-centric trained CNN (above) and location/places-centric trained CNN (below).

Table 3 similarly presents a summary of the top-1 through top-5 predictions for the simulated test mission comprised 16 scenarios of finding persons in specific environmental settings. For these cases, the Theano-AlexNet CNN was retrained, substituting images from the last object class of the ImageNet data set with 952 new images of persons (i.e., a man, woman, boy, or girl) in outdoor scenes. An additional 50 images of persons in outdoor settings were used for validation and 400 additional images were used for the scenario detection testing and analysis.

For the test images of persons outdoors there were 82–215 scenarios of interest detected with no false positives. Figure 5 shows representative neural net model results for the scenarios of finding persons in outdoor scenes.

Table 3 shows that the detection of persons in outdoor settings generally increased as the selection criteria were expanded. A few of the outdoor setting scenarios had a low number of detections (e.g., finding persons on a playground, on a bridge, on a crosswalk, on a railroad track, and in a trench). In these cases, the object-trained CNN appeared to predict class labels for objects in the scene other than a person (e.g., parallel bars, a suspension bridge, a park bench, a maze, a cliff, or a shovel). The object-trained CNN also predicted the class label “military uniform” in addition to or instead of the class label “person” when the test images depicted a Soldier in an outdoor scene.

Table 3 Composite CNN model results for the autonomous detection of persons in specific environmental settings

Simulated test mission scenario	No. of objects recognized					No. of places recognized					No. of scenarios detected					No. of false positives
	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5	1–5
Top-1 through Top-5 criteria																
1. find a person on a beach	7	12	14	15	16	19	23	25	25	25	5	10	14	15	16	0
2. find a person on a boardwalk	9	10	13	16	18	23	24	24	24	24	9	10	13	13	15	0
3. find a person on a bridge	4	7	10	10	10	4	10	10	10	10	2	4	5	5	5	0
4. find a person in a creek	11	13	14	18	19	9	16	16	16	16	3	7	8	11	12	0
5. find a person in a crosswalk	5	7	9	9	11	22	23	23	23	23	5	6	8	8	10	0
6. find a person in a desert	14	18	21	22	22	21	23	23	23	23	10	14	17	18	18	0
7. find a person on a forest path	13	18	20	22	24	16	19	20	21	21	6	12	15	18	20	0
8. find a person in a wheat field	10	12	14	16	17	10	13	14	14	14	3	8	10	12	13	0
9. find a person on a golf course	13	14	18	19	22	24	24	24	24	24	12	13	17	18	21	0
10. find a person on a highway	18	19	20	20	21	9	14	15	16	16	5	10	12	13	14	0
11. find a person on a mountain path	17	19	20	21	21	19	24	25	25	25	11	18	20	21	21	0
12. find a person on a playground	0	0	0	0	0	23	24	24	24	24	0	0	0	0	0	0
13. find a person on a railroad track	2	2	3	4	8	18	18	20	21	21	2	2	3	4	7	0
14. find a person on a ski slope	1	6	10	15	18	23	25	25	25	25	1	6	10	15	18	0
15. find a person in a snow field	11	14	15	16	18	11	19	21	21	21	6	11	13	14	16	0
16. find a person in a trench	5	8	9	11	11	18	22	23	23	23	2	6	8	9	9	0
TOTAL (400 images)	140	179	210	234	256	269	321	332	335	335	82	137	173	194	215	0

OBJECT CNN	person	0.60	person	0.81	person	0.32	person	0.49	person	0.24
	seashore	0.16	hoopskirt	0.02	pier	0.27	bow	0.08	fountain	0.23
	swimming trunks	0.08	pier	0.02	suspension bridge	0.08	rifle	0.02	park bench	0.07
	sandbar	0.05	suspension bridge	0.01	steel arch bridge	0.06	assault rifle	0.02	crutch	0.07
	tank suit	0.05	bannister	0.01	bannister	0.03	reel	0.02	swimming trunks	0.05
ANALYZED IMAGE										
	beach	0.51	boardwalk	0.66	bridge	0.78	creek	0.38	crosswalk	1.00
	coast	0.16	pier	0.20	---	---	river	0.13	---	---
	---	---	---	---	---	---	---	---	---	---
	---	---	---	---	---	---	---	---	---	---
PLACES CNN	person	0.92	person	0.72	person	0.84	person	0.74	person	0.11
	ski	0.02	rapeseed	0.03	hussar monkey	0.02	golf ball	0.11	seashore	0.10
	snowmobile	0.01	hay	0.01	cheetah	0.02	soccer ball	0.02	tandem bicycle	0.08
	dog sled	0.01	sweatshirt	0.01	meerkat	0.01	baseball player	0.02	rickshaw	0.04
	assault rifle	0.01	radio telescope	0.01	chimpanzee	0.01	volley ball	0.02	go-kart	0.04
ANALYZED IMAGE										
	desert sand	0.85	wheat field	0.93	forest path	0.77	golf course	0.93	highway	0.82
	army base	0.13	---	---	---	---	---	---	---	---
	---	---	---	---	---	---	---	---	---	---
	---	---	---	---	---	---	---	---	---	---
OBJECT CNN	person	0.64	person	0.47	person	0.80	person	0.30	person	0.30
	ski	0.07	fountain	0.08	ski	0.08	seashore	0.20	crutch	0.08
	crutch	0.02	bow	0.04	snowmobile	0.03	sand bar	0.13	shovel	0.04
	mountain bike	0.02	ski	0.02	dog sled	0.02	ski	0.04	cliff	0.03
	saluki	0.01	bannister	0.02	ski mask	0.01	dog sled	0.02	coho salmon	0.03
ANALYZED IMAGE										
	mountain path	0.89	railroad track	0.82	ski slope	0.69	snow field	0.96	trench	0.92
	---	---	train station	0.10	snow field	0.15	---	---	---	---
	---	---	---	---	---	---	---	---	---	---
	---	---	---	---	---	---	---	---	---	---
PLACES CNN	---	---	---	---	---	---	---	---	---	---
	---	---	---	---	---	---	---	---	---	---
	---	---	---	---	---	---	---	---	---	---
	---	---	---	---	---	---	---	---	---	---
	---	---	---	---	---	---	---	---	---	---

Fig. 5 Representative composite CNN results for the autonomous detection of the simulated test mission scenario “finding persons in specific environmental settings”. Top-5 image classification labels and top-5 probabilities are annotated for the object-trained CNN (above) and location/places-trained CNN (below). (Top) find a person on the beach; on a boardwalk; on a bridge, in a creek; on a crosswalk. (Middle): find a person in the desert; in a wheat field; on a forest path; on a golf course; on the highway. (Bottom): find a person on a mountain path; on a railroad track; on a ski slope; in a snow field; in a trench.

Table 4 summarizes the top- 1 through top-5 neural predictions for the simulated test mission scenario to find a baseball player playing baseball. Representative test images and CNN model output for this scenario are illustrated in Fig. 6. A total of 16 images were analyzed and 5–11 scenarios of interest were detected with one false positive occurrence related to an image of persons playing sports other than

baseball on a grass field. Figure 7 shows the test image and CNN model predictions for this false positive example, where a baseball player (object) and a baseball field (place) were predicted using the expanded selection criteria; however, the test image depicted football players on a football field.

Ostensibly, more extensive model training using many additional images of tanks and persons in outdoor scenes would be helpful to improve the overall test mission scenario detection outcome, as summarized in Table 5. In Table 5, the percent of detections occurring in the top-1 and top-5 categories are calculated by dividing the number of detections for each category by the number of images analyzed. The last row provides the percent of detections for the entire data set, which is the sum of Army tanks, Persons outdoors, and Baseball player images.

Table 4 Composite CNN model results for the autonomous detection of a baseball player playing baseball

Simulated test mission scenario	No. of objects recognized					No. of places recognized					No. of scenarios detected					No. of false positives				
Top-1 through Top-5 criteria	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
Find a baseball player playing baseball (16)	6	7	8	9	12	7	10	13	13	13	5	6	7	8	11	0	0	1	1	1

DETECTED "BASEBALL GAME" SCENARIO

OBJECT CNN	baseball player	0.93	baseball player	0.59	baseball player	0.91	baseball player	0.96
	baseball	0.07	sundial	0.21	baseball	0.06	baseball	0.32
	---	---	baseball	0.04	hockey puck	0.01	---	---
	---	---	solar dish	0.03	---	---	---	---
	---	---	racket	0.02	---	---	---	---
ANALYZED IMAGE								
	baseball field	0.63	baseball field	0.67	baseball field	0.52	baseball field	0.50
	baseball stadium	0.37	baseball stadium	0.21	baseball stadium	0.48	baseball stadium	0.50
	---	---	outdoor track	0.11	---	---	---	---
	---	---	---	---	---	---	---	---
PLACES CNN	---	---	---	---	---	---	---	---
	---	---	---	---	---	---	---	---
	---	---	---	---	---	---	---	---
	---	---	---	---	---	---	---	---
	---	---	---	---	---	---	---	---

Fig. 6 Composite CNN model results for the autonomous detection of the simulated test mission to find a baseball player playing baseball (i.e., a baseball game scenario). Top-5 image classification labels and top-5 probabilities are annotated for the object-centric trained CNN (above) and location/places-centric trained CNN (below).

OBJECT CNN	football helmet	0.32
	umbrella	0.15
	baseball player	0.06
	balloon	0.05
	parachute	0.04
PLACES CNN	ANALYZED IMAGE	
		
	football stadium	0.58
	baseball field	0.18
	racecourse	0.17
	---	---
	---	---

Fig. 7 Example of the composite CNN model output where there is a false positive occurrence in the detection of a simulated test mission scenario using the extended selection criteria described in the text

Table 5 Summary of the composite CNN model results for the autonomous detection of simulated test mission scenarios

Data set	% of detections occurring in the top-1 and top-5 categories							
	Objects		Places		Scenarios		False positives	
	Top-1	Top-5	Top-1	Top-5	Top-1	Top-5	Top-1	Top-5
Army tanks	55.13	67.95	28.21	62.82	16.67	41.03	0.00	0.00
Persons outdoors	35.00	64.00	67.25	83.75	20.50	53.75	0.00	0.00
Baseball players	37.50	75.00	43.75	81.25	31.25	68.75	0.00	6.25
Sum of Army tanks + Persons outdoors + Baseball players data sets	38.26	64.98	60.32	80.36	20.24	52.22	0.00	0.20

4.1 Cost–Benefit Considerations in Operational Implementation of Neural Network Search Strategies

Operational neural network search strategies should employ a cost–benefit analysis to optimize the modeling system so that detection benefits outweigh the costs (Joo et al. 2003). Generally, the *net benefit* (*nb*) can be expressed as

$$nb = db - cod, \quad (1)$$

where the *detection benefit* (db) are the benefits that accrue from neural network scenario detections. The *cost of detection* (cod) is the cost to resources or time that act to diminish the *net benefit*.

As an example, consider the net benefit of implementing a neural network search strategy, such as the expanded selection criteria described previously, for a hurricane disaster search and rescue mission. In this example, the *net benefit of applying an expanded selection criteria* (nb_{esc}) can be expressed as

$$nb_{esc} = db_{esc} - cod_{fp} . \quad (2)$$

In Eq. 2, the *detection benefit of applying an expanded selection criteria* (db_{esc}) could be the enabled rapid response and dispatch of critical assets to a targeted scene, whereas the *cost of false positives* (cod_{fp}) could be actual negative consequences to both personnel and equipment if unnecessary actions are taken, especially in high-risk or dangerous situations. Incurred *costs* also could be related to the loss of valuable time and/or financial resources. If there were no false positives, then the *detection benefit* of applying an expanded selection criteria to identify more people stranded near roads, on the beach, in rural housing areas, or in heavily vegetated terrain/jungle would need to be weighed against the *cost of false positives* resulting in the operational waste of time, fuel, or personnel resources deployed to deliver needed food or medical supplies to designated locations. Unfortunately, the *cost of false positives* that result in not reaching stranded hurricane victims in the shortest amount of time possible would be the potential loss of lives. The operational implementation of this neural network search strategy would therefore need to be incorporated in such a way as to maximize the *net benefit*. A means to quantify the cost–benefit tradeoff of scenario detection reward versus the false positive cost related risk will need to be further explored in future works.

4.2 Computer Processing, Timing, and Automation

Fast computing speeds are important for realistic autonomous outdoor missions. For the retrained object-centric neural network model described previously, it took approximately 10 s for the code to initialize and then 2 s per test image to produce the top-5 image classification labels and top-5 probabilities. For real-time future implementations of scene understanding software tools, increased computing speeds can be gained by exploiting more efficient algorithms (e.g., Redmon and Farhadi [2017, 2018]), and faster computer processors as they become available. Super computers and parallel processing will enable faster training and validation of neural network codes. Automating the numerical analyses of object- and places-

trained neural network model results would improve future scene understanding software implementations as well. In a compact form factor, improved timing and computer-aided analysis capabilities will allow for real-time implementation of neural network scene understanding software tools in such autonomous assets as robots, drones, and self-driving vehicles.

5. Bringing About Rapid and Robust Autonomous Scene Exploration

In this section, an integrated approach that includes dynamic environmental data retrieval and physics-based modeling is presented, which could augment neural network training to enable enhanced scene understanding and bring about rapid and robust autonomous scene exploration.

5.1 Dynamic Environmental Data Retrieval

Tunick (2016) proposed the incorporation of space and time varying (i.e., dynamic) environmental data from the very beginning of the autonomous data collection process so that the recorded images can be more effectively indexed and retrieved for operational use and analysis. A systematic characterization of the collected images could be useful for improving scene descriptions and help end users develop improved course of action strategies for their autonomous robotic assets. For military as well as emergency services applications, the retrieval of current terrain and weather data can significantly optimize the success of the outdoor mission.

There are many key pieces of information that can be identified as new image data are being recorded that are important and accessible, but are usually overlooked or left undocumented. For example, images can have a timestamp relative to the sun's angle or relative to a world clock. Recorded images can be characterized by the GPS position; the prevailing environmental and weather conditions; and the field of view, depth of view, and image resolution. The dynamic environmental data include the GPS position and altitude above ground level (AGL), prevailing weather, cloud cover, ground and road conditions, and visibility (e.g., fog, smoke, haze, obscurants, or optical turbulence). Identifying environmental and terrain conditions can provide location and geographical context information to help categorize image scenes recorded in diverse regions (e.g., coastal, mountain-valley, desert, forest, urban, rural, ocean, and arctic). Detailed terrain characteristics, such as muddy, sandy, gravelly, wet, dry, or icy, and reports of the most current weather conditions available, such as rain, snow, fog, or haze, can be retrieved and annotated to help describe images used to support the planning or execution of

outdoor tasks. Changing weather conditions, cloud cover, and visibility bring about changes in the illumination of a scene, which can affect image contrast and resolution (Narasimhan and Nayar 2002; Lalonde et al. 2012). Retrieval of time of day and sun angle information is useful to indicate when glare, shadows, or silhouettes may cause difficulties for automated computer vision processes (Shafer 1985; Reddy and Veeraraghavan 2014). Taking note of optical turbulence conditions is important because these effects can significantly degrade and blur image quality due to spatial smearing (Roggemann et al. 1996).

Camera specifications and the image data measurements themselves are also important (e.g., the spatial and temporal image resolutions, field of view, depth of view, and scene color or shading variations). Together with the environmental information, these camera/image elements can provide further details for scene description and image indexing. In most cases, the image/camera information can be annotated based on the camera type, lensing, pixel array, and timing specifications. For example, co-located range finder instrumentation could provide effective depth and field-of-view measurements for this purpose.

When communications are available, the most current environmental information available can be extracted from several accessible resources. Obtaining the most current data available is advantageous since environmental conditions (e.g., weather and terrain) can change over very short temporal and spatial intervals. Access to data from Department of Defense (DOD) GPS can provide latitude and longitude or Universal Transverse Mercator (UTM) location and timestamp information, commonly reported as Greenwich Mean Time (GMT) or Coordinated Universal Time (UTC). Data from the US Naval Observatory (USNO) can provide precise timing information as well as solar and lunar elevation/azimuth angles. Terrain and geographical location and context information are provided by satellite and aerial imagery for military operations from the US Army Corps of Engineers, Army Geospatial Center or from public Internet resources such as Google, MapQuest, Bing, and Yahoo Maps.

Weather conditions and related oceanic, atmospheric, and geophysical data are available for the military through the US Air Force (USAF) 557th Weather Wing (i.e., formerly the USAF Weather Agency) and for the civilian community through the National Weather Service (NWS) and the National Centers for Environmental Information. Daily NWS weather reports that are found online contain hourly records citing the date, time, wind speed (miles per hour), visibility (miles), weather (i.e., rain, snow, fog, haze, and so on), sky/cloud condition (reported as overcast [OVC], broken [BRK], scattered [SCT] or clear [CLR] along with the cloud ceiling height in hundreds of feet AGL), air temperature, dew point temperature, relative

humidity (%), pressure, and precipitation (in inches). Current weather and weather forecast information are readily found on Internet websites, such as Intellicast, AccuWeather, and Weather Underground. In areas where communications are either restricted or unavailable, the information needed to describe the scene should instead be retrieved from co-located sensors on the robots themselves or gleaned from the recorded images using neural network models trained on dynamic environment features.

5.2 Physics-Based Modeling

People are innately capable of making rapid physical inferences about objects or perceived activities in a scene and answering the question “what happens next?” (Battaglia et al. 2013; Ullman et al. 2014; Wu et al. 2015; Lake et al. 2016). For computational scene understanding, such inferences can be achieved by combining neural network training with physics-based predictive modeling to help predict how objects in a scene interact with their surroundings. For instance, physics-based modeling can predict how the physical properties of the terrain (as affected by the weather) will impact the navigation of a robotic system. Zhang et al. (2016) discussed two types of physical scene understanding models. They conducted a study to compare “intuitive physics engines” with “memory-based models”. In this case, physics simulation engines approximate how objects in complex scenes interact under the laws of physics over short time periods (e.g., stability analysis for stacks of blocks). In contrast, memory-based neural networks make predictions of outcomes in a new scene based on “stored experiences” of encountered scenes and physical outcomes. Wu et al. (2015) presented a model for detecting physical properties of objects by integrating a physics engine with deep learning neural network predictions. Similarly, Fragkiadaki et al. (2016) reported on how an autonomous agent could be equipped with an internal predictive physics model of the surrounding environment, and how one could use the physics-based model in combination with neural network training to predict actions that previously have not been encountered by the agent. The combined approach of Fragkiadaki et al. (2016) was demonstrated by accurately predicting the required actions for an autonomous agent to play a simulated billiards game.

These physics-based models may seem complicated and difficult to implement. However, there are alternate types of physics-based models for terrain and weather forecasting already available (Chenery 1997; McDonald et al. 2016; HQDA 1989, 2015) that can be implemented directly on a robotic system or accessed via reach back network communications. One can also envision using weather forecast models together with terrain and morphology data to set up physics-based mission

rehearsals and training. Furthermore, where terrain and weather data are available, they can be incorporated using nowcast techniques to interpolate and reconstruct local information needed to support the autonomous outdoor mission.

5.3 Integrated Scene Understanding

To better illustrate the advantages of an integrated scene understanding approach, where data assimilation along robotic paths and physical modeling are added to neural networks, an instructive analysis of the representative real-world, time-sensitive missions listed in Table 1 is presented.

5.3.1 Ground Search and Rescue

Ground search and rescue is an autonomous outdoor scene exploration mission that involves terrain and obstacle traversal, acoustic detection of endangered personnel, as well as robotic lifting of heavy objects (e.g., people and/or debris and rubble from fallen buildings) (Levinger et al. 2008). Neural network training that specifically addresses changing environmental dynamics can provide useful clues related to ground conditions, visibility, and illumination to benefit navigation and robotic search algorithms. Adding environmental data retrieval and physics-based weather forecasting can provide analyses of current and changing conditions (e.g., those that may affect traversability for robot route planning in uneven terrain). With regard to robots operating over extended periods of time, it would be important to consider changes in scene illumination from day to night and changes in the visual appearance of a scene due to changes in weather and the seasons (Pepperell et al. 2014; Sunderhauf et al. 2015; Neubert and Protzel 2016).

5.3.2 Aerial Reconnaissance

The aerial-reconnaissance autonomous outdoor scene exploration mission involves the takeoff, landing, and in-flight control of aerial autonomous assets as well as obstacle avoidance and autonomous use of visible and IR imaging sensors (HQDA 2009; Korpela 2016; Levin et al. 2016). Real-time dynamic environmental data retrieval combined with physics-based weather forecast modeling can augment neural network training by providing critical updates on atmospheric stability conditions essential for aerial missions. Terrain and morphology data retrieval can be used to augment both navigation and areal search models. Environmental data such as temperature, air density, cloud cover, wind speed, wind direction, visibility, and precipitation can be used to augment physics-based optical turbulence modeling to assess refractive index effects on aerial image quality (Roggemann et al. 1996).

5.3.3 Nuclear, Biological, and Chemical (NBC) Hazard Detection

The NBC hazard detection autonomous outdoor scene exploration mission strives to identify and quantify NBC hazards. It involves robotic NBC sensors, real-time air-quality data retrieval, and NBC hazard analysis (HQDA 1986; Scott 2003). Clearly, real-time turbulence and atmospheric stability data are vital for downwind NBC hazard prediction, to include identifying toxicity levels and assessing the pervasiveness and persistence of the threat. Integrating current and forecasted weather elements is equally important because in combat, for example, the weather can alter terrain features and traversability; low visibility can impede reconnaissance or alternately conceal friendly forces' maneuvers and activities; and wind speed and direction can favor upwind personnel in the event of an NBC attack or decrease the effectiveness of downwind personnel and equipment due to blowing dust, smoke, sand, rain, or snow. NBC hazard detection missions also can be those associated with chemical spills from overturned trucks or trains or from accidents at industrial plants. In this situation, neural network training on objects, places, and changing environmental features can provide many critical components for the scene understanding image analysis.

5.3.4 Cave and Tunnel Reconnaissance

The cave and tunnel reconnaissance autonomous scene exploration mission supports the exploration of underground spaces and facilities. It involves underground mapping, terrain and obstacle traversal, laser illumination, and visible and IR imaging (Magnuson 2013; Eshel 2014). Neural network training can provide image classification of salient and meaningful features related to open or concealed objects as well as recognition of subterranean morphology and ground cover conditions. Physics-based models could help predict the outcome and mission impact of relevant robot, terrain, and morphology interactions (e.g., those associated with traversability). In a potentially communications-limited region, retrieval of co-located temperature, pressure, and humidity sensor data can supplement physics-based models, for instance, to assess microclimate effects, such as condensation or frost on robotic vision systems.

5.3.5 Sniper Detection

The sniper detection autonomous outdoor scene exploration mission searches for difficult to find weaponized threats. It involves acoustic detection of gunfire, laser illumination of identified signal/source locations, and visible and IR imaging to identify potential threats from a distance (Crane 2006). In this situation, neural network training can provide person, object, building, and place recognition capabilities as well as identifying vital weather and terrain features, such as ground

conditions for navigation and visibility for reconnaissance. Physics-based acoustic and optical models can provide analysis of propagation conditions for optimal sensor and equipment performance. Physics-based weather forecast modeling and terrain assessment can predict whether ground or visibility conditions will change if the weather worsens during operations. Concurrently, environmental data retrieval, physics-based modeling, and neural network training can provide enhanced situational awareness to support the mission.

5.4 Applicability to Diverse Evolving and Time-Sensitive Missions

The analysis we provided can be applied to diverse evolving and time-sensitive missions in a manner similar to those summarized in Table 6. The mission of ground search and rescue can be applied not only to downed pilots but to lost campers and children. Aerial reconnaissance could apply not only to airplanes but also to drones for military and domestic exploration (e.g., via visible and IR imaging [Levin et al. 2016]). NBC hazard detection not only applies to downwind hazards in warzones but also to chemical fires from overturned trucks and trains or from industrial warehouses. The approach to cave and tunnel reconnaissance can be adapted to finding lost explorers, looking for infiltrating terrorists, or looking for victims of subway or deep mining accidents. Unfortunately, sniper detection is needed for both military and domestic situations to respond to menacing threats, whether on the battlefield or in urban and rural areas. There can be many other outdoor missions where the need for rapid and robust autonomous scene exploration applies, such as fighting forest fires, ensuring boating and harbor safety, and monitoring air-pollution hazards.

Table 6 Integrated scene understanding approach for five representative real-world, time-sensitive missions (summary)

Real-world, time sensitive mission	Neural network training				Physics-based models										Dynamic environmental data retrieval							
	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t		
Ground search and rescue																						
Terrain and obstacle traversal		x	x							x				x	x	x						
Autonomous scene exploration via visible and IR imaging	x	x	x	x			x	x	x					x	x	x						
Acoustic detection					x									x	x	x						
Lifting heavy objects (e.g., people, rubble, and debris)	x		x								x			x								
Aerial reconnaissance																						
Takeoff, landing, and in-flight control			x	x					x	x		x		x	x	x						
Obstacle avoidance	x	x	x						x	x		x		x		x						
Autonomous scene exploration via visible and IR imaging	x	x	x	x			x	x	x					x	x	x						
NBC hazard detection																						
NBC sensors data retrieval									x	x								x	x	x		
NBC hazard analysis									x				x	x	x	x		x	x	x		
Autonomous scene exploration via visible and IR imaging	x	x	x	x			x	x	x				x	x	x							
Cave and tunnel reconnaissance																						
Terrain and obstacle traversal			x							x				x			x					
Laser illumination						x								x			x					
Autonomous scene exploration via visible and IR imaging	x		x	x				x						x			x	x				
Sniper detection																						
Acoustic detection					x									x	x	x						
Laser illumination						x	x							x	x	x						
Autonomous scene exploration via visible and IR imaging	x	x	x	x			x	x	x					x	x	x						

Key:

- | | | |
|--|---|---|
| a. Open/concealed object recognition | i. Weather/terrain forecasts | q. Robot weather sensors: co-located temperature, pressure, and humidity in communications-denied region |
| b. Building/location identification | j. Robot/terrain/morphology interaction | r. Air-quality measurements |
| c. Terrain features/ground cover recognition | k. Payload stability analysis | s. NBC sensor measurements |
| d. Weather feature identification | l. Aerodynamics | t. Atmospheric turbulence/stability |
| e. Acoustic propagation, refraction, and scattering | m. NBC downwind hazard analysis | |
| f. Optical propagation, refraction, and scattering | n. Local surface and lower atmosphere wind, temperature, pressure, air density, and humidity | |
| g. Optical turbulence | o. Regional scale weather trends | |
| h. Image enhancement/turbulence mitigation | p. GPS/timestamp/ sun angle | |

6. Rapid and Robust Neural Detection Scheme Models

Neural scene understanding for autonomous outdoor missions requires robust recognition of objects and places in realistic environmental settings. Neural detection speed will aid in the effectiveness of neural models to support robotic time sensitive and temporally evolving missions. While there are a number of computer models for neural scene understanding (Krizhevsky et al. 2012; Zhou et al. 2014, 2017; Karpathy and Li 2015; Lim et al. 2017; Qiao et al. 2017; Yu et al. 2017; Yang et al. 2017; Han et al. 2017), the CNN-yolov3 model was reported to be fast and able to work on a variety of media formats (Redmon et al., 2016; Redmon and Farhadi, 2017, 2018). In one of our setups, model calculations took approximately 0.03 s to produce a single output image frame. The CNN-yolov3 model implements a novel object detection scheme to produce both object label and bounding box neural predictions, which may be useful for some autonomous outdoor missions. Proof-of-principle experiments were conducted using image data recorded on a robotics platform at the US Army Research Laboratory (ARL) to evaluate the suitability of the CNN-yolov3 model for rapid and robust neural detection to support robotic missions.

To determine how well the CNN-yolov3 model can detect a specified image target in an outdoor scene, test image data were collected using a camera mounted on a robotics platform at ARL (Fig. 8). Recorded test image data were used to produce the CNN-yolov3 model result shown in Fig. 9, illustrating that a slightly visible person standing behind tree branches was found in the image. For this example, the image target (person) was standing at a distance of about 30 m from the robot-mounted camera. A panoramic view of this model result is shown in Fig. 10. Note that in Fig. 10, the robot tracks are visible in the near field.



Fig. 8 Robotic platform with mounted camera at ARL



Fig. 9 Robotic neural detection of a person behind tree branches. The test image was obtained from video recorded on a robotics platform at ARL.

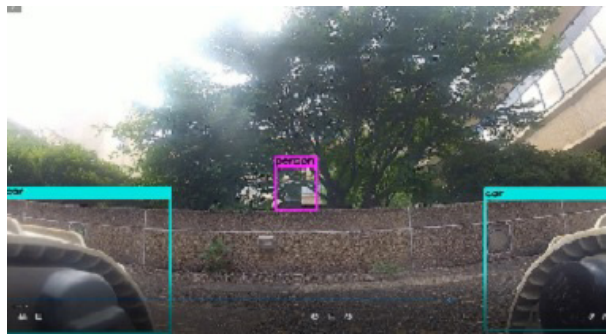


Fig. 10 Panoramic robotic neural detection of a person behind tree branches. Note the robot track in the near-field. The test image was obtained from video recorded on a robotics platform at ARL.

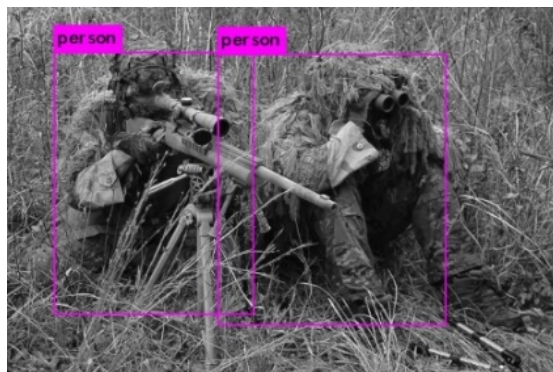


Fig. 11 Neural detection of persons (Soldiers) in camouflage. The test image was obtained from the Internet.

Besides images collected from a robot, the CNN-yolov3 model also was evaluated using publically available images. Representative CNN-yolov3 model results from these tests are shown in Figs. 11–14. The results shown in Figs. 11–14 demonstrate that the CNN-yolov3 model is capable of detecting one or more persons in a scene. For autonomous robotic missions in realistic environments, our proof-of-principle experiments show that available neural detection schemes can achieve sufficient speed and appear to be useful for detecting some objects in outdoor scenes, such as persons, but that more work is needed to achieve robustness.



Fig. 12 Neural detection of a person (Soldier) in camouflage and vegetative terrain. The test image was obtained from the Internet.



Fig. 13 Neural detection of a person (Soldier) in vegetative terrain. The test image was obtained from the Internet.

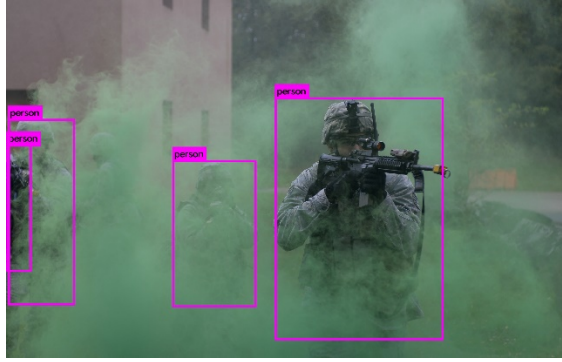


Fig. 14 Neural detection of persons (Soldiers) in obscurants. The test image was obtained from the Internet.

6.1 Autonomous Inference of Mission Intent

The autonomous inference of mission intent adds a fourth dimension to neural scene understanding, both by recognizing and potentially controlling evolving outcomes of detected and surveilled activities. Foremost among potential autonomous outdoor missions where intelligent robots can be useful are those dealing with security. Combatants augmented with adversarial lethal robots are an emerging threat that could be countered with the aid of autonomous outdoor missions. To counter this emerging threat, autonomous detection and reasoning tools need to be developed that are capable of inferring mission intent, in particular, adversarial mission intent. Once the adversarial mission intent is understood, a counter strategy can be implemented. If the surveilled mission intent is benign, then a counter strategy may not be needed.

The detection of objects in realistic environmental settings can provide clues to infer mission intent. For example, a neural detection scheme may need to distinguish between robots augmenting noncombatants, friendlies, or enemy combatants with different uniforms, insignias, weapons, and equipment, and in different locations. However, neural models designed to infer mission intent need to do more than detect objects and places. Neural detection models need to be augmented with algorithms to enable autonomous inference and reasoning about the scene content and evolution in space and time. As an example, the neural detection of chem-bio agent weaponized robots augmenting enemy combatants wearing helmets and gas masks may be inferred as the adversary's intent to release airborne toxic agents and launch a chem-bio attack. The autonomous inference of mission intent could be useful in other settings also, such as detecting friend or foe or detecting an insider threat. We have shown that autonomous inference of mission intent can be useful to support security operations and counter evolving threats,

particularly with the development of improved autonomous detection and reasoning tools.

7. Conclusions

A novel proof-of-principle neural network software configuration trained on objects and places data sets was shown to have applicability to a wide class of autonomous detection problems. Representative scenarios for three simulated test missions were identified by autonomously detecting environmental settings in addition to objects in the scene. Proof-of-principle experiments showed that a composite CNN trained on separate image databases can be adapted to the detection of real-world, time-sensitive scenarios that are keyed to the success of the outdoor mission. Analysis of five representative real-world missions indicated that adding environmental data assimilated along robotic paths and physical modeling to neural networks would further improve neural scene understanding and benefit the autonomous detection of key time-sensitive and evolving scenarios. Developing capabilities for the autonomous inference of mission intent would enable even more neural detection benefits by both recognizing and potentially controlling evolving outcomes of detected and surveilled activities.

8. References

- Aminoff EM, Tarr MJ. Associative processing is inherent in scene perception. *PLOS One*. 2015;10(6):e0128840.
- Aminoff EM, Toneva M, Shrivastava A, Chen X, Misra I, Gupta A, Tarr MJ. Applying artificial vision models to human scene understanding. *Front Comput Neurosci*. 2015;9(8):1–14.
- [ARL] US Army Research Laboratory S&T Campaign Plans 2015–2035: System Intelligence and Intelligent Systems, 98–99. Army Research Laboratory (US); 2014.
- [AWC] Army Transformation Wargame 2001. Carlisle Barracks (PA): Army War College (US); 2001 Apr 22–27.
- Bar M. Visual objects in context. *Nat Rev Neurosci*. 2004;5(8):617–29.
- Baro X, Gonzalez J, Fabian J, Bautista MA, Oliu M, Jair Escalante H, Guyon I, Escalera S. Chalearn looking at people 2015 challenges: Action spotting and cultural event recognition. *IEEE Conference on Computer Vision and Pattern Recognition, Workshop on ChaLearn Looking at People*; 2015. p. 1–9.
- Battaglia PW, Hamrick JB, Tenenbaum JB. Simulation as an engine of physical scene understanding. *PNAS*. 2013;110(45):18327–18332.
- Biederman I. Recognition-by-components: A theory of human image interpretation. *Psychol Rev*. 1987;94:115–148.
- Blei DM, Ng AY, Jordan MI. Latent Dirichlet allocation. *J Mach Learn Res*. 2003;3:993–1022.
- Chenery JT. The terrain model: A miniature battlefield. *MIPB*. 1997;23(4).
- Crane D. Anti-sniper/sniper detection/ gunfire detection systems at a glance. *Defense Review*; 2006 July 16 [accessed 2018 Mar 12]. <http://defensereview.com/anti-snipersniper-detectiongunfire-detection-systems-at-a-glance/>.
- Daniels ZA, Metaxas D. Scenarios: a new representation for complex scene understanding. 2018 Feb 16. arXiv:1802.06117.
- Ding W, Wang R, Mao F, Taylor G. Theano-based large-scale visual recognition with multiple GPUs. 2015. arXiv:1412.2302v4.

- Eshel T. Operating robots underground. Defense Update; 2014 July 27 [accessed 2017 June 14]. http://defense-update.com/20140727_underground-robots.html.
- Fragkiadaki K, Agrawal P, Levine S, Malik J. Learning visual predictive models of physics for playing billiards. 2016. arXiv:1511.07404v3.
- Gallant J, David S. The neural prediction challenge. Berkley (CA): UC Regents; 2015 [accessed 2018]. <http://neuralprediction.berkeley.edu>.
- Gori I, Aggarwal JK, Matthies L, Ryoo MS. Multi-type activity recognition in robot-centric scenarios. *EEE Robot Autom Let*. 2016;1(1): 593–600.
- Hall LT, Beart GCG, Thomas EA, Simpson DA, McGuinness LP, Cole JH, et al. High spatial and temporal resolution wide-field imaging of neuron activity using quantum NV-diamond. *Sci Rep*. 2012. <https://doi.org/10.1038/srep00401>.
- Han X, Zhong Y, Cao L, Zhang L. Pre-trained AlexNet architecture with pyramid pooling and supervision for high spatial resolution remote sensing image scene classification. *Remote Sens*. 2017;9:848.
- [HQDA] Field behavior of NBC agents (including smoke and incendiaries), Appendix C. Washington (DC): Headquarters, Department of the Army (US); 1986. Field Manual No.: FM 3–6. Appendix C written by Meyers RE.
- [HQDA] Weather support for Army tactical operations. Washington (DC): Headquarters, Department of the Army; 1989. Field Manual No.: FM 34-81.
- [HQDA] Army unmanned aircraft system operations. Washington (DC): Headquarters, Department of the Army; 2009. Field Manual No.: FM 3-04.155.
- [HQDA] Weather support and services for the US Army. Washington (DC): Headquarters, Department of the Army; 2015. Field Manual No.: AR 115-10.
- Howard N, Cambria E. Intention awareness: Improving upon situation awareness in human-centric environments. *HCIS*. 2013;3(1);1–17.
- Itti L, Koch C, Niebur E. A model of saliency-based visual attention for rapid scene analysis. *IEEE Trans Pattern Anal Mach Intel*. 1998;20(11):1254–1259.
- Joo D, Hong T, Han I. The neural network models for IDS based on the asymmetric costs of false negative errors and false positive errors. *Expert Sys Appl*. 2003 [https://doi.org/10.1016/S0957-4174\(03\)00007-1](https://doi.org/10.1016/S0957-4174(03)00007-1).

- Judson J. US Army putting finishing touches on autonomous systems strategy. Defense News. 2016 Mar 17.
- Karadas M, Wojciechowski AM, Huck A, Dalby NO, Andersen UL, Thielscher A. Opto-magnetic imaging of neural network activity in brain slices at high resolution using color centers in diamond. 2017. arxiv:arXiv:1708.06158v1.
- Karpathy A, Li FF. Deep visual-semantic alignments for generating image descriptions. IEEE Conference on Computer Vision and Pattern Recognition; 2015; Boston, MA.
- Korpela C, Root P, Kim J, Wilkerson S, Gadsden SA. A framework for autonomous and continuous aerial intelligence, surveillance, and reconnaissance operations. Proc SPIE 9828, Airborne Intelligence, Surveillance, Reconnaissance (ISR) Systems and Applications XIII; 2016 May 17; Baltimore, MD.
- Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. Proc. 25th International Conference on Advances in Neural Information Processing Systems; 2012.
- Lake BM, Ullman TD, Tenenbaum JB, Gershman SJ. Building machines that learn and think like people. Behav Brain Sci. 2016. <https://doi.org/10.1017/S0140525X16001837>.
- Lalonde JF, Efros AA, Narasimhan SG. Estimating the natural illumination conditions from a single outdoor image. Int J Comput Vis. 2012;98:123–145.
- Levin E, Zarnowski A, McCarty JL, Bialas J, Banaszek A, Banaszek S. Feasibility study of inexpensive thermal sensors and small UAS deployment for living human detection in rescue missions application scenarios. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences; 2016. <https://doi.org/10.5194/isprsarchives-XLI-B8-99-2016>.
- Levinger J, Hofmann A, Theobald D. Semiautonomous control of an emergency response robot. Proc Association for Unmanned Vehicle Systems International (AUVSI) Unmanned Systems North America Conference; 2008 June 10–12; San Diego, CA.
- Li L, Li F. What, where and who? Classifying events by scene and object recognition. IEEE 11th International Conference on Computer Vision; 2007. <https://doi.org/10.1109/ICCV.2007.4408872>.

- Lim K, Hong Y, Choi Y, Byun H. Real-time traffic sign recognition based on a general purpose GPU and deep-learning. PLOS ONE. 2017;12(3):e0173317.
- Magnuson S. Reconnaissance robots' place on battlefields still unsettled. National Defense; 2013 Oct 31 [accessed 2018 Mar 12]. <http://www.nationaldefensemagazine.org/articles/2013/10/1/2013october-reconnaissance-robots-place-on-battlefields-still-unsettled>.
- Malcolm GL, Groen II, Baker CI. Making sense of real-world scenes. Cog Sci. 2016;20(11):843–856.
- McDonald EV, Bacon SN, Bassett SD, Amit R, Enzel Y, Minor TB, McGwire K, Crouvi O, Nahmias Y. Integrated terrain forecasting for military operations in deserts: Geologic basis for rapid predictive mapping of soils and terrain features. In: McDonald EV, Bullard T, editors. Military Geosciences and Desert Warfare. New York (NY): Springer; 2016. p. 353–375.
- Narasimhan SG, Nayar SK. Vision and the atmosphere. Int J Comput Vis. 2002;48(3):233–254.
- Neubert P, Protzel P. Beyond holistic descriptors, keypoints, and fixed patches: Multiscale superpixel grids for place recognition in changing environments. IEEE Robot Autom Lett. 2016;1(1).
- Patterson G, Hays J. Sun attribute database: Discovering, annotating, and recognizing scene attributes. Proc. IEEE Conference on Computer Vision and Pattern Recognition; 2012; Providence, RI.
- Pepperell E, Corke, PI, Milford MJ. All-environment visual place recognition with SMART. Proc. IEEE International Conference on Robotics and Automation; 2014 May 31–June 7; Hong Kong, China.
- Perazzi F, Krahenbuhl P, Pritch Y, Hornung A. Saliency filters: contrast based filtering for salient region detection. Proc. IEEE Conference on Computer Vision and Pattern Recognition; 2012 June 16–21; Providence, RI.
- Piekiewicz F, Laurent P, Petre C, Richert M, Fisher D, Hylton T. Unsupervised learning from continuous video in a scalable predictive recurrent network. 2016. arXiv:1607.06854v3.
- Qiao K, Chen J, Wang L, Zeng L, Yan B. A top-down manner-based DCNN architecture for semantic image segmentation. PLOS ONE. 2017;12(3):e0174508.

- Reddy D, Veeraraghavan A. Lens flare and lens glare. In: Ikeuchi K, editor. In: Computer vision: a reference guide. New York (NY): Springer; 2014. p. 445–447.
- Redmon J, Divvala S, Girshick R, Farhadi A. You only look once: unified, real-time object detection. IEEE Conference on Computer Vision and Pattern Recognition, 2016 June 26–July 1; Las Vegas, NV.
- Redmon J, Farhadi A. Yolo9000: Better, faster, stronger. Proc. IEEE Conference on Computer Vision and Pattern Recognition; 2017; Honolulu, HI. p. 6517–6525.
- Redmon J, Farhadi A. YOLOv3: An incremental improvement. 2018. arXiv:1804.02767v1.
- Roberts R, Ta DN, Straub J, Ok K, Dellaert F. Saliency detection and model-based tracking: a two part vision system for small robot navigation in forested environment. Proc. SPIE 8387, Unmanned Systems Technology XIV Conference; 2012.
- Roggemann MC, Welsh BM, Hunt BR. Imaging through turbulence. Boca Raton (FL): CRC Press; 1996.
- [RCTA] Robotics Collaborative Technology Alliance. FY 2012 Annual Program Plan. Adelphi (MD); Army Research Laboratory (US); 2012.
- Russakovsky O, Deng J, Su H, Krause J, Satheesh S, Ma S, Huang Z, Karpathy A, Khosla A, Bernstein M, et al. ImageNet large scale visual recognition challenge. Int J Comput Vis. 2015;115(3):211–252.
- Scott AJ, Mabesa JR, Hughes C, Gorsich DJ, Auner GW. Equipping small robotic platforms with highly sensitive more accurate nuclear, biological, and chemical (NBC) detection systems. Proc SPIE 5083, Unmanned Ground Vehicle Technology V; 2003.
- Shafer SA. Shadows and silhouettes in computer vision. Berlin (Germany): Kluwer Academic Publishers; 1985.
- Sauve G, Harmand M, Vanni L, Brodeur MB. The probability of object–scene co-occurrence influences object identification processes. Exp Brain Res. 2017;1–13.
- Sofer I, Crouzet SM, Serre T. Explaining the timing of natural scene understanding with a computational model of perceptual categorization. PLOS Comput Biol. 2015;11(9):e1004456.

- Sunderhauf N, Shirazi S, Dayoub F, Upcroft B, Milford M. On the performance of ConvNet features for place recognition. Proc IEEE International Conference on Intelligent Robots and Systems; 2015 Sept. 28–Oct. 2; Hamburg, Germany.
- Sustersic J, Wyble B, Advani S, Narayanan V. Towards a unified multiresolution vision model for autonomous ground robots. Rob Auton Sys. 2016;75:221–232.
- Tai L, Liu M. Deep-learning in mobile robotics - from perception to control systems: a survey on why and why not. 2017. arXiv:1612.07139v3.
- Tunick A. Space–time environmental image information for scene understanding. Adelphi (MD): Army Research Laboratory (US); 2016. Report No.: ARL-TR-7652.
- Ullman, TD, Stuhlmuller A, Goodman ND, Tenenbaum JB. Learning physics from dynamical scenes. Proc. 36th Annual Conference of the Cognitive Science Society; 2014 Jul 23–26; Quebec City, Canada.
- Valada A, Oliveira G, Brox T, Burgard W. Towards robust semantic segmentation using deep fusion. Robotics: Science and Systems (RSS 2016), Workshop, Are the Sceptics Right? Limits and Potentials of Deep Learning in Robotics; 2016.
- Wang L, Wang Z, Du W, Qiao Y. Object-scene convolutional neural networks for event recognition in images. IEEE Conference on Computer Vision and Pattern Recognition, Workshop on ChaLearn Looking at People; 2015. p. 30–35.
- Wang L, Wang Z, Qiao Y, Van Gool L. Transferring deep object and scene representations for event recognition in still images. Int J Comput Vis. 2017. <https://doi.org/10.1007/s11263-017-1043-5>.
- Warnell G, David P, Chellappa R. Ray saliency: bottom-up visual saliency for a rotating and zooming camera. Int J Comput Vis. 2016;116(2):174–189.
- Wohler C, Anlauf JK. Real-time object recognition on image sequences with the adaptable time delay neural network algorithm - applications for autonomous vehicles. Image and Vis Comput. 2001;19:593–618.
- Wu J, Yildirim I, Lim JJ, Freeman B, Tenenbaum J. Galileo: Perceiving physical object properties by integrating a physics engine with deep learning. Adv Neural Inf Process Sys. 2015;127–135.

- Xiang T, Gong S, Parkinson D. Autonomous visual events detection and classification without explicit object-centred segmentation and tracking. In: Proc British Machine Vision Conference; 2002. p. 21.1–21.10.
- Xiao J, Hays J, Ehinger K, Oliva A, Torralba A. SUN database: large-scale scene recognition from abbey to zoo. IEEE Conference on Computer Vision and Pattern Recognition; 2010;Jun 13–18; San Francisco, CA.
- Xu J, Yang Z, Tsien JZ. Emergence of visual saliency from natural scenes via context-mediated probability distributions coding. PLOS ONE. 2010;5(12):e15796.
- Yang S. Hazardous chemical dispersion analysis and surrogate model optimization using artificial neural network. [Doctorate thesis]. [Republic of Korea]: The Graduate School of Seoul National University; 2017.
- Yang Z, Jiang W, Xu B, Zhu Q, Jiang S, Huang W. A convolutional neural network-based 3D semantic labeling method for ALS point clouds. Rem Sens. 2017;9:936.
- Yeomans B, Shaukarm A, Gaom Y. Testing saliency based techniques for planetary surface scene analysis. Proc. 13th Symposium on Advanced Space Technologies in Robotics and Automation; 2015 May 11–13; Noordwijk, the Netherlands.
- You Q, Luo J, Jin H, Yang J. Robust image sentiment analysis using progressively trained and domain transferred deep networks. 2015. arXiv:1509.06041v1.
- Yu S, Wu S, Wang L, Jiang F, Xie Y, Li L. A shallow convolutional neural network for blind image sharpness assessment. PLOS ONE. 2017;12(5):e0176632.
- Zhang R, Wu J, Zhang C, Freeman WT, Tenenbaum JB. A comparative evaluation of approximate probabilistic simulation and deep neural networks as accounts of human physical scene understanding. 2016. arXiv:1605.01138v2.
- Zhou B, Lapedriza A, Xiao J, Torralba A, Oliva A. Learning deep features for scene recognition using Places database. Advances in Neural Information Processing Systems 27; 2014 Dec 8–13; Montreal, Canada.
- Zhou B, Lapedriza A, Khosla A, Oliva A, Torralba A. Places: a 10 million image database for scene recognition. IEEE Trans Pattern Anal Mach Intell. 2017a;40(6):1452–1464. <https://doi.org/10.1109/TPAMI.2017.2723009>.

Zhou B, Lapedriza A, Khosla A, Torralba A, Oliva A. Places demo. Cambridge (MA): Massachusetts Institute of Technology; 2017b [accessed 2018]. <http://places2.csail.mit.edu/demo.html>.

List of Symbols, Abbreviations, and Acronyms

AGL	above ground level
ARL	US Army Research Laboratory
BRK	broken
CNN	convolutional neural network
CLR	clear
cod	cost of detection
cod_{fp}	cost of false positives
db	detection benefit
db_{esc}	expanded selection criteria
DOD	Department of Defense
GMT	Greenwich Mean Time
GPS	global positioning system
ILSVRC2012	ImageNet Large Scale Visual Recognition Challenge 2012
IR	infrared
nb	net benefit
NBC	Nuclear, Biological, and Chemical
nb_{esc}	net benefit of applying an expanded selection criteria
NWS	National Weather Service
OVC	overcast
SCT	scattered
USAF	US Air Force
USNO	US Naval Observatory
UTC	Universal Time
UTM	Universal Transverse Mercator

1 DEFENSE TECHNICAL
(PDF) INFORMATION CTR
DTIC OCA

2 DIR ARL
(PDF) IMAL HRA
RECORDS MGMT
RDRL DCL
TECH LIB

1 GOVT PRINTG OFC
(PDF) A MALHOTRA

7 ARL
(PDF) RDRL CII
S YOUNG
RDRL CII A
D BARAN
A TUNICK
RDRL CII B
S RUSSELL
RDRL CIN T
R DROST
R E MEYERS
K DEACON