ARL-TR-9198 ● May 2021

DEVCOM
ARMY RESEARCH
LABORATORY

Synthetic Environments for Artificial Intelligence (AI) and Machine Learning (ML) in Multi-Domain Operations

by Raghuveer Rao, Celso de Melo, and Hamid Krim

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.

ARL-TR-9198 ● MAY 2021

Synthetic Environments for Artificial Intelligence (AI) and Machine Learning (ML) in Multi-Domain Operations

Raghuveer Rao and Celso de Melo
Computational and Information Sciences Directorate,
DEVCOM Army Research Laboratory

Hamid Krim
Army Research Office, DEVCOM Army Research Laboratory

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*
May 2021

2. REPORT TYPE
Technical Report

3. DATES COVERED (From - To)
8–9 December 2020

4. TITLE AND SUBTITLE
Synthetic Environments for Artificial Intelligence (AI) and Machine Learning (ML) in Multi-Domain Operations

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Raghuveer Rao, Celso de Melo, and Hamid Krim

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
DEVCOM Army Research Laboratory
ATTN: FCDD-RLC-CI
Adelphi, MD 20783-1138

8. PERFORMING ORGANIZATION REPORT NUMBER
ARL-TR-9198

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 unlimited.

13. SUPPLEMENTARY NOTES
ORCID IDs: Raghuveer Rao, 0000-0001-6481-4175, Celso de Melo, 0000-0003-2680-8334, Hamid Krim, 0000-0003-4971-1690

14. ABSTRACT

The use of artificial intelligence solutions for Army field applications will rely heavily on machine learning (ML) algorithms. Current ML algorithms need large amounts of mission-relevant training data to enable them to perform well in tasks such as object and activity recognition and high-level decision-making. Battlefield data sources can be heterogeneous, encompassing multiple sensing modalities. Present open-source data sets for training ML approaches provide inadequate representation of scenes and situations of interest to the Army, in terms of both content and sensing modalities. There is a push to use synthetic data to make up for the paucity of real-world training data relevant to military multi-domain operations of the future. However, there are no systematic approaches for synthetic generation of data that provide any degree of assurance of improved real-world performance of the ML techniques trained on such data. The problem of effective synthetic data generation for ML raises deeper questions than that of artificially generating speech or imagery that humans find realistic. An Army Science Planning and Strategy Meeting held in December 2020 explored multiple technical issues in depth related to synthetic data generation and application to Army problems of interest.

15. SUBJECT TERMS
artificial intelligence, machine learning, synthetic data, multi-domain operations, Army sciences, military information sciences

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON**
					Raghuveer Rao
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**	UU	22	**19b. TELEPHONE NUMBER (Include area code)**
Unclassified	Unclassified	Unclassified			(301) 394-0860

Contents

List of Figures

Acknowledgments

We thank Dr Patrick Baker and Dr Alex Kott for approving the Army Science Planning and Strategy Meeting (ASPSM) on Synthetic Training Environments for Artificial Intelligence and Machine Learning in Multi-Domain Operations. Dr Philip Perconti's early support for investigating synthetic data as an important topic in the context of training machine-learning algorithms has been key to organizing this ASPSM. We thank the invited speakers whose preparation and presentation were instrumental to the ASPSM's success: Prof Antonio Torralba (Massachusetts Institute of Technology [MIT]), Prof James DiCarlo (MIT), Prof Jitendra Malik (University of California, Berkeley [UC Berkeley]), Prof Leonidas Guibas (Stanford University), Prof Jessica Hodgins (Carnegie Mellon University), Prof Trevor Darrell (UC Berkeley), Prof Rama Chellappa (Johns Hopkins University), Prof Judy Hoffman (Georgia Institute of Technology [Georgia Tech]), and Prof Kate Saenko (Boston University). Stimulating and lively discussions were possible because of our panelists: Dr Piotr Franaszczuk (US Army Combat Capabilities Development Command [DEVCOM] Army Research Laboratory), Prof Dhruv Batra (Georgia Tech), Prof Dieter Fox (University of Washington), Dr Behzad Kamgar-Parsi (Office of Naval Research), Prof Amir Shirkhodaie (Tennessee State University), Prof Alex Dimakis (University of Texas at Austin), Dr Joseph Reynolds (DEVCOM Command, Control, Communications, Computers, Cyber, Intelligence, Surveillance and Reconnaissance), Dr Yajie Zhao (University of Southern California Institute for Creative Technologies [USC-ICT]), Prof Mani Srivastava (University of California, Los Angeles), and Dr Mohammad Soleymani (USC-ICT). Many others contributed to the organization in multiple ways, from managerial support to the conduct of the sessions: Ms Tamara Christenson, Mr Greg Lieberman, Mr Andrew Ladas, Dr Purush Iyer, Dr Lance Kaplan, Dr Brent Lance, Dr Tien Pham, Dr Edward Palazzolo, Dr Robert St. Amant, Dr Sean Hu, Mr Matthew Thielke, Dr Heesung Kwon, Ms Danielle Caruccio, Dr Suya You, Dr Jim Uplinger, Dr Priya Narayanan, Mr Patrick Rauss, Dr Susan Young, Dr Hyungtae Lee, and Dr Sungmin Eum. We thank all the participants for their contribution to the many productive discussions in the ASPSM.

1. Introduction

Artificial intelligence (AI) is a modernization priority for US defense. The DOD AI strategy directs the department to accelerate the adoption of AI and the creation of a force fit for our time.[1] It is natural then that it is also an Army modernization priority.[2] It is an important element of solving problems from the Army's perspective of multi-domain operations (MDO) as built on layered standoff in adversary engagements.[3] While there is no concise and universally accepted definition of AI itself, the DOD AI strategy document[1] refers to it as, "the ability of machines to perform tasks that normally require human intelligence – for example, recognizing patterns, learning from experience, drawing conclusions, making predictions, or taking action – whether digitally or as the smart software behind autonomous physical systems". This statement implies a machine is exhibiting intelligence when it performs such tasks on its own, independently of human assistance. A key aspect of AI solutions that have emerged in the last decade is that they overwhelmingly fit the pattern-recognition mold; in most cases, they are assigning input data to data classes based on the outputs of a trained artificial neural network (ANN) to the same input data. Specifically, deep-learning neural networks (DNNs), consisting of many layers of artificial neurons and connecting weights, are initially trained on large amounts of data from known classes to determine the weights and then used to classify actual input data in an application.[*] Thus machine learning (ML), the process by which automata, in this case DNNs, learn patterns in the training phase, has been a dominant theme. Indeed, the success of DNNs in computer vision has been responsible for the increased attention and investment in AI by the commercial and government sectors.[4] Advances in training algorithms and software development tools such as tensorflow, availability of computational power such as graphical processing units (GPUs), and access to large amounts of data such as through social media have enabled rapid exploration of deep-learning models in many applications.

In supervised learning, where human experts create a set of samples to train the ML algorithms, the closeness of the training data to actual application data plays an important role in the AI approach's performance. The main bottleneck for application of ML models to military problems is the lack of representative data in sufficient volume to train these models.[5] The use of synthetic data has been proposed as a workaround.[6] Synthetic data sets offer certain advantages:

[*] The number of weights, or parameters, can be huge, in the millions. This puts a tremendous burden on computation and training time.

- They come with accurate ground-truth.
- It is easy to generate various types of data in large volumes with off-the-shelf simulation products.
- They impose fewer procedural hurdles such as, for example, obtaining institutional review board permissions with biometric data.

However, the most crucial issue is whether training the ML models on synthetic data, or mixed synthetic and real data, enables these models to perform well with real data. Initial results obtained by the US Army Combat Capabilities Development Command Army Research Laboratory researchers and collaborators using synthetically generated human videos for robot recognition of gestures, show that training on a mix of synthetic and real data can improve the performance of the ML gesture recognizer.[7] However, there are no universal or categorical results indicating consistent improvement of real-world ML performance when trained, fully or partially, on synthetic data. A systematic investigation is therefore necessary to determine the degree of confidence with which one can employ synthetic data for training ML methods. It is reasonable to hypothesize that the effectiveness of synthetic data for improving ML performance will be influenced, among other factors, by the domain of actual application, the fidelity of the synthesized data to real data, the training regime, and the ML methods themselves. Fidelity of synthesized data to real data, in turn, depends on the data synthesis methods and raises the issue of assessing the fidelity through suitable metrics. It is not clear, with images for example, that the performance of ML methods with synthetic data training is proportionately related to their fidelity with real scenes as perceived by human vision. It is possible that there are key features of the data that are more important for ML performance than those that influence human perception. A key purpose of organizing this Army Science Planning and Strategy Meeting (ASPSM) was to have leading academic and DOD experts in synthetic data generation, artificial intelligence and machine learning (AI & ML), and human perception address these very issues. The technical emphasis of the meeting was predominantly on image and video data reflecting the organizers' mission areas of computer vision and scene perception.

2. Organization

Based on the issues raised in the previous section, the meeting was organized around three topical thrusts:

1. *Human learning and generalization:* Humans can generalize from minimal abstractions and descriptions to complex objects. For example, observing a

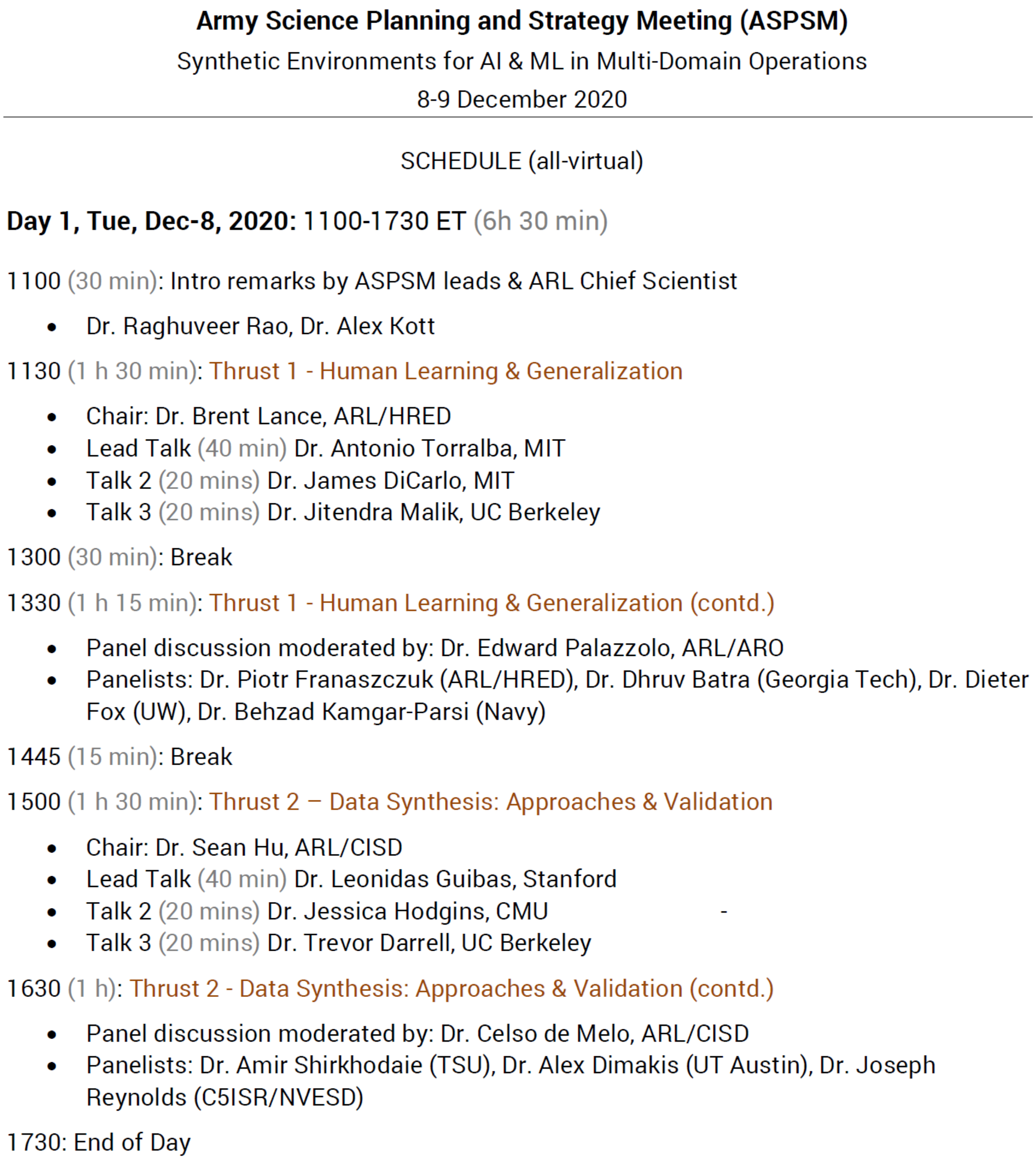
cartoon image or line drawing of an object would suffice in many situations for humans to recognize the actual 3-D object later in a real scene in spite of the latter possessing more complex attributes than the cartoon or drawing depiction.[*] This is way beyond the ability of current AI & ML systems. If such capability were to be developed, it would significantly reduce the burden on the data-synthesis machine to assure tight fidelity across all attributes of the real data. This example is also an illustration of the fact that research into synthetic data generation for training ML models is intricately connected with improving the capabilities of the ML models themselves. This thrust was thus focused on exploration of learning in humans and animals to inspire new approaches in ML and data synthesis.

2. *Data synthesis approaches and validation:* Most areas in which ML methods are being applied have techniques and tools available for synthesizing data specific to their domains. Gaming platforms provide a popular commercial example of video synthesis. There is the question of how to evaluate the performance of the different synthesis approaches in a given domain. One clearly has to identify metrics or criteria to perform such evaluation. Typically, too, authors of synthesis tools issue claims on the performance or efficacy of the tools. Validation would be the process of assessing such claims. The intent of this thrust was to address principles governing both the synthesis and validation processes. Examples of synthesis techniques of interest include computer graphics-based renderers (e.g., as used in movies), physics-based simulation (e.g., IR imagery), and generative models (which currently tend to be neural network based).

3. *Domain adaptation challenges:* Domain adaptation in ML refers to training the ML model using data from one domain, known as the source, and then applying the ML to data in a different but related domain, known as the target.[9] An example would be that of an ML algorithm trained to recognize vehicles using a source-image data set of predominantly civilian vehicles and then using the trained algorithm to recognize vehicles in a target data set containing mostly military vehicles. Where synthetic data are used for training, they would typically constitute the source domain and actual application data would be the target domain. The focus of this thrust was to identify and discuss key issues and challenges in effective domain adaptation.

[*] Closely related issues are discussed by Lamb et al.[8]

The ASPSM deliberations were spread over four sessions. Day 1 had two sessions covering the first two topical thrusts. The first session of Day 2 covered the third thrust and the second session was devoted to breakout discussions under the three thrusts. The schedules for the two days of the ASPSM are shown in Figs. 1 and 2, respectively. As can be seen, each thrust-based session began with a 40-min lead talk by an academic expert in the field followed by two 20-min talks, again by university experts. The talks were followed by discussions with a panel consisting of a mix of experts from academia and DOD. The last session consisted of breakouts where participants could discuss various aspects related to the thrusts.



Army Science Planning and Strategy Meeting (ASPSM)

Synthetic Environments for AI & ML in Multi-Domain Operations

8-9 December 2020

SCHEDULE (all-virtual)

Day 1, Tue, Dec-8, 2020: 1100-1730 ET (6h 30 min)

1100 (30 min): Intro remarks by ASPSM leads & ARL Chief Scientist

- Dr. Raghuveer Rao, Dr. Alex Kott

1130 (1 h 30 min): Thrust 1 - Human Learning & Generalization

- Chair: Dr. Brent Lance, ARL/HRED
- Lead Talk (40 min) Dr. Antonio Torralba, MIT
- Talk 2 (20 mins) Dr. James DiCarlo, MIT
- Talk 3 (20 mins) Dr. Jitendra Malik, UC Berkeley

1300 (30 min): Break

1330 (1 h 15 min): Thrust 1 - Human Learning & Generalization (contd.)

- Panel discussion moderated by: Dr. Edward Palazzolo, ARL/ARO
- Panelists: Dr. Piotr Franaszczuk (ARL/HRED), Dr. Dhruv Batra (Georgia Tech), Dr. Dieter Fox (UW), Dr. Behzad Kamgar-Parsi (Navy)

1445 (15 min): Break

1500 (1 h 30 min): Thrust 2 – Data Synthesis: Approaches & Validation

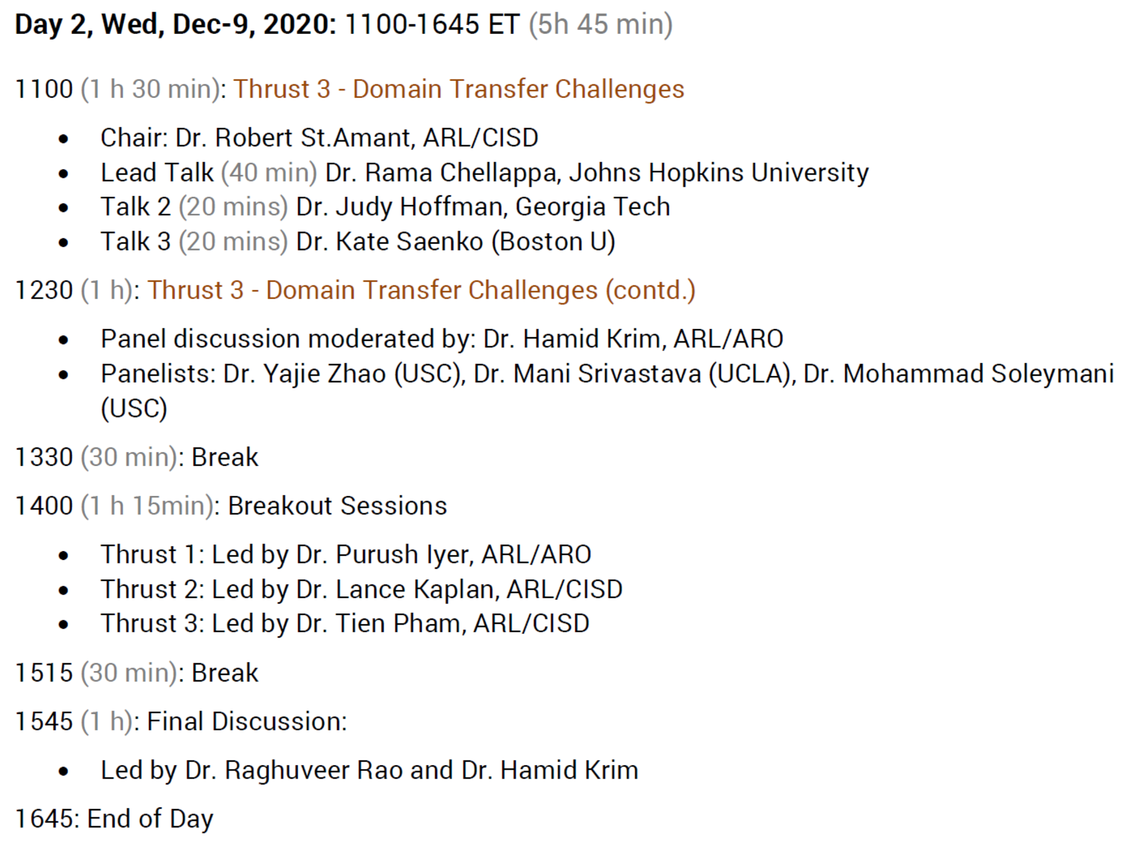
- Chair: Dr. Sean Hu, ARL/CISD
- Lead Talk (40 min) Dr. Leonidas Guibas, Stanford
- Talk 2 (20 mins) Dr. Jessica Hodgins, CMU -
- Talk 3 (20 mins) Dr. Trevor Darrell, UC Berkeley

1630 (1 h): Thrust 2 - Data Synthesis: Approaches & Validation (contd.)

- Panel discussion moderated by: Dr. Celso de Melo, ARL/CISD
- Panelists: Dr. Amir Shirkhodaie (TSU), Dr. Alex Dimakis (UT Austin), Dr. Joseph Reynolds (C5ISR/NVESD)

1730: End of Day

Fig. 1 Day 1 schedule



Day 2, Wed, Dec-9, 2020: 1100-1645 ET (5h 45 min)

1100 (1 h 30 min): Thrust 3 - Domain Transfer Challenges

- Chair: Dr. Robert St.Amant, ARL/CISD
- Lead Talk (40 min) Dr. Rama Chellappa, Johns Hopkins University
- Talk 2 (20 mins) Dr. Judy Hoffman, Georgia Tech
- Talk 3 (20 mins) Dr. Kate Saenko (Boston U)

1230 (1 h): Thrust 3 - Domain Transfer Challenges (contd.)

- Panel discussion moderated by: Dr. Hamid Krim, ARL/ARO
- Panelists: Dr. Yajie Zhao (USC), Dr. Mani Srivastava (UCLA), Dr. Mohammad Soleymani (USC)

1330 (30 min): Break

1400 (1 h 15min): Breakout Sessions

- Thrust 1: Led by Dr. Purush Iyer, ARL/ARO
- Thrust 2: Led by Dr. Lance Kaplan, ARL/CISD
- Thrust 3: Led by Dr. Tien Pham, ARL/CISD

1515 (30 min): Break

1545 (1 h): Final Discussion:

- Led by Dr. Raghuveer Rao and Dr. Hamid Krim

1645: End of Day

Fig. 2 Day 2 schedule

3. Oral Sessions and Panels

Prof Antonio Torralba of the Electrical Engineering & Computer Science Department of MIT delivered the lead talk in Session 1 on human learning and generalization. It was titled "Learning from vision, touch and audition" and it provided insights on how deep-learning approaches can discover meaningful representations of scenes without the use of extensive labeled training data. The illustrated example involved their DNN developing associations between visual scenery and sounds in the environment. The reader is referred to Aytar et al.[10] for a representative article on this topic.

The next talk by Dr James DiCarlo, also from MIT, was titled "Reverse Engineering Visual Intelligence". The speaker defined reverse engineering as inferring internal processes in the brain based on observations of behaviors and reactions to inputs, and forward engineering as creating ANN models that would generate corresponding behaviors with the same inputs. A goal of his group is that of establishing benchmarks of performance of neurocognitive tasks that would simultaneously be met by humans, or other primates, and by ML models.[11] His talk presented initial results showing how models of processing in the brain could be adapted to ANN implementation and made the case that ANNs, which emulate human behavior closely through incorporation of these adaptations, would in turn be accurate descriptions of brain functioning.

The third talk in Session 1, by Prof Jitendra Malik of UC Berkeley, was titled “Turing’s Baby”. The title is perhaps a reference to one of the earliest electronic stored-program computers, nicknamed “Baby”, one of whose creators was inspired by Alan Turing.[12] Prof Malik began by quoting Turing’s musing that, rather than creating a program to simulate the adult mind, one could begin by simulating that of a child.[13] Essentially, this would mean creating an AI that would learn and grow by interacting with the environment and by learning from other AI and humans. This is referred to as embodied machine intelligence. Prof Malik argued that supervised learning essentially deals with static data sets and thus displays disembodied intelligence operating on a well-curated moment in time. Specifically, he posited that the supervisory training approach is ill-suited for creating ML that can provide human-level understanding of the world, especially of human actions. Prof Malik presented “Habitat”, a platform developed by him and his collaborators for research in embodied AI.[14] The panel discussion that followed touched on the topics addressed by the speakers as well as those related to learning by robots, and current models of intelligence development in children.

Session 2 on “Data Synthesis: Approaches and Validation” began with a talk, titled “Learning to Generate or Generating to Learn?” by Prof Leonidas Guibas of Stanford University. Among the motivations for investigating synthetic data generation for training ML, he pointed out the alleviation of the burden of massive amounts of human annotations of training data. His premise was that generation efficiency and realism of the synthetic data are important regardless of whether used for training ML or for human consumption. However, he indicated other metrics of quality are not well defined and need further study. He showed examples of improved object-recognition performance by ML when trained on mixed synthetic and real data but acknowledged the difficulty of drawing generalizable conclusions.

Dr Jessica Hodgins of Carnegie Mellon University delivered the second talk of the second session, titled “Generating and Using Synthetic Data for Training”. The talk showed examples of finely detailed synthetic scenes generated by her research group. Using the process of style transfer from real scenes to synthetic scenes,[15] her group has created examples of where ML methods, trained on preponderantly style-adapted synthetic data mixed with some real data, have outperformed those trained just on either real-only or synthetic-only data sets. The explanation for the improved performance is attributed to the style transfer overcoming the “distribution gap” between the synthetic and real data sets.

The final talk of Session 2 was by Prof Trevor Darrell of UC Berkeley. Titled “Generating, Augmenting, and Adapting Complex Scenes”, it was divided into three parts. The first part detailed a technique developed by the speaker and his co-researchers, called “semantic bottleneck scene generation”, for synthesizing scenes from ground-truth labels. The techniques could further be combined with models generating such ground labels through a generative process. Detailed description of the technique is available in Azadi et al.[16] The second part dealt with augmentation and self-supervised learning. The speaker made the argument that current contrastive learning methods build invariances in synthesizing augmented data that may or may not be beneficial. For example, building rotational invariance might be beneficial in recognizing flowers in a scene but might hinder effective recognition of objects with specific orientations. The speaker described his group’s approach of considering multiple learning paths with specific invariances and showed results indicating improved performance over prior art.[17] The third part presented a technique called “Tent” (for “Test Entropy”). The premise is that the data encountered during application of a DNN might be distributed differently from the training data, resulting in performance degradation. Real-time or test-time adaptation of DNN parameters is therefore desirable to prevent such degradation. The Tent technique achieves this by minimizing the measured entropy of the DNN output by adapting its weights. The speaker then showed improved performance of this technique over prior approaches with commonly used data sets.[18] The ensuing panel discussion dealt with synthesis challenges, especially those of IR images.

Day 2 began with Session 3 on “Domain Transfer Challenges”. Dr Rama Chellappa, a Bloomberg Distinguished Professor at Johns Hopkins University, gave the first talk, titled “On the Expectations and Maximization of Synthetic Data for Solving Real DOD Problems”. The talk began by tracing the history of multiple DOD programs dealing with synthetic imagery over the last two decades. It made a key assertion that domain shift between real and synthetic data may be less if the physics governing the real data is incorporated in the synthesis process. Prof Chellappa also provided a quick tutorial on domain adaptive representations covering formal mathematical approaches as well as the newer generative adversarial networks (GANs). The GAN-based approach developed by the speaker and his coresearchers modifies the distribution of synthetic data to match that of the target distribution. The talk showed examples of this approach outperforming prior non-GAN approaches.[19]

Prof Judy Hoffman of the Georgia Institute of Technology delivered the next talk, titled “Challenges in Generalizing from Multiple Data Sources”. The problem she considered is that of learning models in simulation that then transfer to the real world. She identified four challenges: Generation, Enumeration, Generalization, and Adaptation. The speaker presented several different approaches for addressing these challenges. Specifically, the domain-specific masks for generalization (DMG) approach tackles multisource domain learning by balancing domain-specific with domain-invariant feature representations to produce a single model that provides effective domain generalization.[20]

The third and final talk of Session 3, titled "Recent Advances and Challenges in Sim2Real Domain Transfer for Image Classification and Segmentation", was given by Prof Kate Saenko of Boston University. While continuing on themes addressed in the previous two talks, Prof Saenko also provided a history of visual domain adaptation and addressed issues of domain and data-set biases. Among different approaches for correcting data-set bias, the talk dealt in detail with domain adaptation. Of particular significance was the ability of techniques developed by Prof Saenko and collaborators showing synthetic to real adaptation, as from game engines to real data.[21] The ensuing panel discussion brought up several interesting questions including that of training and test domains being different, not in the objects of interest but in the environments the objects are found in, such as military vehicles in a desert environment during training but in a tropical-vegetation background during testing.

4. Breakout Discussions

The three breakout discussions for each of the three thrusts were conducted in parallel. Discussions in the breakout on "Human Learning & Generalization" began by addressing questions such as, "How do humans learn?", "How do ML models mimic human processes?", and "How do synthetic data enable these?" Relationships between learning and growth from childhood through adolescence and adulthood emerged as key points. Other factors identified as aspects of human learning that would help if transferred to ML were human psychology, emotions, simultaneous engagement in multidimensional activities, memory, and ability to unlearn.

The breakout on "Data synthesis: Approaches & Validation" identified several issues with synthesizing data, especially with image and video. The main questions related to the usefulness of incorporating physics, tradeoffs between fidelity to visual appearance and cost, metrics for fidelity, the importance of fidelity itself, and limitations of current techniques including those of GANs. It was observed that synthetic image and video generation has been around for at least a couple of decades but most products were designed either for visual effects or for reproducing physical measurements (e.g., radiometric profiles in IR simulation). They are not well suited for training ML. Another issue raised was the importance of synthesized 2-D imagery to be consistent with the underlying 3-D geometry of objects and environment. The case was also made that being able to generate a large amount of synthetic data in a particular context of interest may serve to test new AI & ML approaches as a first pass, regardless of whether that results in such methods working well on real data.

The Thrust 3, “Domain Transfer Challenges”, breakout discussion identified a key desired AI capability for MDO as that of going from isolated learning to joint or collaborative learning among machines and humans. Joint learning in the sense of training ML simultaneously on multiple modalities of data was also discussed. It was recognized that work has barely begun in these areas. The need for providing unambiguous specification to the Soldier of what an AI-based system will do in a given situation was emphasized by the breakout lead. This led to discussions on system robustness. The breakout leads provided a summary of the discussions to the ASPSM audience.

5. Gaps and Recommendations

Based on the deliberations of this ASPSM, we have identified the following as areas worthy of further Army science and technology (S&T) investment:

1. *Synthesis techniques and data sets that support multimodal interactive learning*. In contrast to the prevailing static data sets that capture “moments in time” (e.g., images of vehicles in rural settings), it is necessary to develop simulators that are more representative of the embodied experiences that support continuous learning, as we see with humans, and enable richer representations of the world. Hybrid approaches (e.g., augmented reality) may also bring together the advantages of human supervision with the flexibility of synthetic environments.

2. *Algorithms and architectures to learn and synthesize causality and hierarchical relationships.* Recent approaches, such as graph-based convolutional neural networks, have shown promise in learning hierarchical relations in space and time (e.g., object-part and cause-effect relationships). Given the complexity of collecting and annotating such data in the real-world, synthetic data generation could be particularly useful. Identifying hierarchical relationships is a key ingredient of general DOD and battlefield intelligence analysis.

3. *Algorithms and architectures that support continuous, incremental, multimodal learning.* Deep-reinforcement-learning methods are being successfully used to train virtual or robotic agents on relevant action policies such as predator–prey interactions. Imitation-based approaches acknowledge the social aspect of learning and typically partner agents with (often human) teachers to learn new policies. These types of interactive continuous learning can further be paired with multimodal learning (i.e., fusing data from multiple sensors) to enable richer representations of the world that are more robust and generalizable. Again, the difficulty of obtaining large amounts of curated data in this realm provides motivation for exploring synthesis engines.

4. *Algorithms and architectures that learn physics or are endowed with relevant physics domain knowledge.* In many domains (e.g., object perception in IR light), perceiving from images and synthesizing imagery requires an understanding of the underlying physical properties of the world such as interaction between light and material. However, current deep-learning models lack this physical knowledge. Developing techniques that endow ML with physics domain knowledge is critical to the performance of these systems.

5. *Domain adaptation techniques with rich intermediate representations.* To close the domain gap between real and synthetic data, it is essential to further current trends in building domain-invariant intermediate representations, in particular, using semantic dictionaries and generative adversarial networks. Representations that are able to understand the underlying structure of the data (e.g., lighting, rotation, color) are more likely to succeed in abstracting away from unimportant details in the synthetic data.

6. *Methods for providing insight into ML models' internal representations and comparison of synthetic versus real representations.* Network dissection techniques "open up" the hidden layers in deep-learning models, allowing interpretation of which particular concepts, or their finer aspects, are being learned at each stage in the network. These techniques shed light on the internal representations of DNNs with real and synthetic inputs, helping identify key differences in what is being learned and, consequently, finding solutions to overcome such differences.

6. Conclusions

The two-day virtual ASPSM drew a large and enthusiastic attendance of DOD scientists and engineers, leading academic experts, and S&T program managers. The multiple multidisciplinary discussions reinforced the view that developing improved methods for generating synthetic data for training ML approaches cannot be separated from understanding and improving the ML approaches themselves. A particularly important need is that of understanding how ML approaches, especially current learning architectures, create an internal representation of the scene. Two other areas that emerged as important are 1) understanding similarities and differences between human learning and what is possible in the ML world, and 2) multimodal data—from both synthesis and ML perspectives. We anticipate increased collaborative efforts in the near term between DOD and academic researchers in the areas identified in this report.

7. References

1. US Department of Defense. Summary of the 2018 Department of Defense artificial intelligence strategy: Harnessing AI to advance our security and prosperity. US Department of Defense; 2018.

2. Headquarters, Department of the Army. Army modernization strategy: Investing in the future. Headquarters, Department of the Army; 2019.

3. US Army Training Doctrine and Command. Pamphlet TP525-3-1, the US Army in multi-domain operations 2028. US Army Training Doctrine and Command; 2018.

4. Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. In: Proceedings of the 25th International Conference on Neural Information Processing Systems; vol. 1; 2012; Red Hook, NY.

5. Allen G. Understanding AI technology. Joint Artificial Intelligence Center. 2020 [accessed 2021 Apr 13]. https://apps.dtic.mil/sti/citations/AD1099286.

6. Narayanan P, Borel-Donohue C, Lee H, Kwon H, Rao RM. A real-time object detection framework for aerial imagery using deep neural networks and synthetic training images. In: Signal Processing, Sensor/Information Fusion, and Target Recognition XXVII. Proceedings of SPIE DCS; 2018.

7. de Melo C, Rothrock B, Gurram P, Ulutan O, Manjunath BS. Vision-based gesture recognition in human-robot teams using synthetic data. Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS); 2020; Las Vegas, NV.

8. Lamb A, Ozair S, Verma V, Ha D. Sketchtransfer: A challenging new task for exploring detail-invariance and the abstractions learned by deep networks. Proceedings of the IEEE Winter Conference on Applications of Computer Vision (WACV); 2020; Snowmass Village, CO.

9. Patel VM, Gopalan R, Li R, Chellappa R. Visual domain adaptation: a survey of recent advances. IEEE Signal Processing Magazine. 2015 May;32(3):53–69.

10. Aytar Y, Vondrick C, Torralba A. SoundNet: Learning sound representations from unlabeled video. Proceedings of the 30th International Conference on Neural Information Processing Systems (NIPS ’16); 2016; Red Hook, NY.

11. Schrimpf M, Kubilius J, Lee MJ, Murty NAR, Ajemian R, DiCarlo JJ. Integrative benchmarking to advance neurally mechanistic models of human intelligence. Neuron. 2020;108(3):413–423.

12. Copland B. The manchester computer: A revised history part 2: the baby computer. IEEE Annals of the History of Computing. 2011;33(1)22–37.

13. Turing AM. Computing machinery and intelligence. Mind. 1950;LIX(236).

14. Savva M, Kadian A, Maksymets O, Zhao Y, Wijmans E, Jain B, Straub J, Liu J, Koltun V, Malik J, Parikh D, Batra D. Habitat: A platform for embodied AI research. Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV); 2019; Seoul, South Korea.

15. Gatys LA, Ecker AS, Bethge M. Image style transfer using convolutional neural networks. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 2016; Las Vegas, NV.

16. Azadi S, Tschannen M, Tzeng E, Gelly S, Darrell T, Lucic M. Semantic bottleneck scene veneration. 2019 [accessed 2021 Apr 14]. https://arxiv.org/abs/1911.11357.

17. Xiao T, Wang X, Efros AA, Darrell T. What should not be contrastive in contrastive learning. 2021 [accessed 2021 Apr 14]. https://arxiv.org/abs/2008.05659v2.

18. Wang D, Shelhamer E, Liu S, Olshausen B, Darrell T. Tent: Fully test-time adaptation by entropy minimization. 2021 [accessed 2021 Apr 14]. https://arxiv.org/abs/2006.10726v3.

19. Sankaranarayanan S, Balaji Y, Jain A, Lim SN, Chellappa R. Learning from synthetic data: Addressing domain shift for semantic segmentation. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR); 2018; Salt Lake City, UT.

20. Chattopadhyay P, Balaji Y, Hoffman J. Learning to balance specificity and invariance for in and out of domain generalization. In: Vedaldi A, Bischof H, Brox T, Frahm JM, editors. Computer vision. Proceedings of ECCV 2020. Springer Cham; 2020.

21. Hoffman J, Tzeng E, Park T, Zhu J-Y, Isola P, Saenko K, Efros A, Darrell T. CyCADA: Cycle-consistent adversarial domain adaptation. Proceedings of the 35th International Conference on Machine Learning; 2018.

List of Symbols, Abbreviations, and Acronyms

2-D	two-dimensional
3-D	three-dimensional
AI	artificial intelligence
AI & ML	artificial intelligence and machine learning
ANN	artificial neural network
ASPSM	Army Science Planning and Strategy Meeting
DMG	domain-specific masks for generalization
DNN	deep-learning neural network
DOD	Department of Defense
GAN	generative adversarial network
GPU	graphical processing unit
IR	infrared
MDO	multi-domain operations
ML	machine learning
S&T	science and technology

1 (PDF)	DEFENSE TECHNICAL INFORMATION CTR DTIC OCA
1 (PDF)	DEVCOM ARL FCDD RLD DCI TECH LIB
3 (PDF)	DEVCOM ARL FCDD RLC CI R RAO C DE MELO H KRIM