

"Ducunt volentem fata, nolentem trahunt."

Fate leads the willing and drags the unwilling.
—Seneca

"Drive Change ... Forge Victory."
—LTG Milford H. Beagle, 2024

Introduction

Recent advances in the fields of artificial intelligence (AI), computer vision and convolutional neural networks have begun impacting the wider world in highly visible ways. For example, knowledge generation applications such as Perplexity AI, Bing Copilot and Phind facilitate accurate text and image

response to a wide variety of user prompts. In this way, by leveraging a form of inductive reasoning, nascent AI chatbots interact with specific user prompts and formulate tailored responses. It would seem as if these applications were on the verge of sensemaking or at least something remarkably close to it.

Contemporary writers have observed this phenomenon and applied it to future AI applications within the military domain. For example, Warrant Officer 1 (WO1) Clifford A. Baxt illustrates how AI can optimize the sensor-to-shooter chain. Similarly, Norine MacDonald and George Howell discuss how unmanned aerial vehicles of all sizes could use increasingly accurate forms of machine vision for potential target classification. Their ideas provide excellent initial observations into how future AI applications can optimize targeting. However, all three writers stay within the realm of general explanation. In other words, they provide a superficial explanation of how umbrella terms such as AI, machine learning and computer vision may facilitate future targeting.

In part, this article intends to build upon their ideas while providing the specific explanation of how Deep Convolutional Neural Networks (DCNNs) can facilitate automatic target acquisition for the fires warfighting function. Essentially, this article will get into the specifics of how to build AI applications for the purpose of warfighting. Consequently, I argue, mature applications of DCNNs will play an outsized role in future conflict because the side with the highest quality DCNN will be able to more rapidly find, classify and target adversarial combat power formations.

Deep Convolutional Neural Networks

This section portrays a general sense of what DCNNs are, how they are trained, how they operate and how they are currently performing in the contemporary military domain. In this way, the intent is to familiarize the reader with this novel technology. Once a general sense of DCNNs has been achieved, the article will move more fully into the military domain while explaining how DCNNs will contribute to future targeting cycles within the U.S. fires warfighting function.

To start at the beginning, think of DCNNs as a way to reach a desired end; that end is what theorists refer to as artificial general intelligence or AGI. Microsoft researchers Sebastien Bubeck et al. "use AGI to refer to systems that demonstrate broad capabilities of intelligence, including reasoning, planning and the ability to learn from experience, and with these capabilities at or above human level." Therefore, DCNNs are the stacked neural networks that enable a machine to understand, learn and, most importantly, remember things. Memory is what facilitates extended learning. Keeping this in mind, DCNNs are an imperfect plastic representation of the human brain.

Haohan Wang and Bhiksha Raj trace the origin of DCNNs all the way back to 300 BC when Aristotle introduced what contemporary researchers refer to as associationism. Associationism was Aristotle's method for understanding how the human mind learns and remembers. For example, in his book Memory and Reminiscence, Aristotle asserts that the human mind recalls data and experience through four laws: the laws of (1) contiguity, (2) contrast, (3) frequency and (4) similarity.⁴ The law of contiguity refers to recalling memories that may be "spatially joined but essentially different." Therefore, these are memories of different things that occurred in the same time period. The law of contrast refers to the opposite of similarity or recollections that are defined in opposition to other memories. The law of frequency encapsulates memories which an individual finds him or herself continuously pondering. And lastly, the law of similarity refers to memories that share common characteristics. According to Aristotle, these are the four methods a human brain leverages to learn to recall memory.

Keeping this information in mind, DCNNs use the latter three laws (contrast, frequency and similarity) during supervised, unsupervised and semi-supervised training for the purpose of correctly classifying an object within a bounded box in the real-world. This ability to classify objects in the real world is referred to as computer vision as well as object detection. Computer vision and object

detection is what allows a myriad of machines, such as unmanned aerial systems, to participate in automatic target acquisition. However, to reach a relative level of competence, a machine must first be trained.

DCNNs are trained in one of three ways: supervised, unsupervised and semi-supervised. Supervised training refers to DCNNs being fed labeled data by a human supervisor so the DCNN algorithm can then build a "predictive model" which it can then recall memory for the purpose of classifying objects in the real world. For example, a DCNN that has been constructed to recognize Russian-built S-300 air defense platforms will be fed thousands, millions or perhaps even billions of different S-300 images. The similarity and frequency of the S-300 pictures construct what the DCNN will recognize as an S-300. Furthermore, as the DCNN is fed different labeled images of other types of Air Defense Artillery (ADA) systems, it will learn to differentiate or contrast between the different types of systems. In this way, supervised training constructs a DCNN algorithm which has been specifically designed to classify an image after receiving some sort of input data, usually in the form of still picture or video.

Conversely, unsupervised training refers to DCNNs being initially fed non-labeled data. Unsupervised training still facilitates object detection, only in a different manner. For example, unsupervised training researchers state:

"The DCNN architecture is designed such that the network learns automatically the 'important' underlying pattern of the data. Therefore, [researchers] can train DCNNs to learn the features using unlabeled data. This is called feature learning. Then, after training the DCNN models in [an] unsupervised way, they are used to extract features from a small amount of labeled data which are used to train classifiers in a supervised way." ¹⁰

Therefore, an unsupervised training method enables the machine to learn on its own while self-correcting during the latter stages of training with a small amount of labeled data.

Sticking with the S-300 example, a DCNN would be specifically constructed for S-300 object identification. Then, while training in an unsupervised manner, the DCNN would engage with unlabeled data for the purpose of differentiating between S-300 images and non-S-300 images. After its training session is complete, referred to as an epoch, the DCNN compares its identifications with a small set of labeled S-300 images. Utilizing the Aristotelian law of similarity, unsupervised training still produces the machine's ability to detect and classify images.

Lastly, semi-supervised training refers to DCNNs which undergo training epochs that use both labeled and unlabeled data simultaneously. This training method treats unlabeled data as variables that the machine must "iterate" to accurately categorize along with the labeled data. Referring back to our S-300 example, a semi-supervised training epoch would feed the DCNN labeled S-300 images alongside unlabeled S-300 images and other random images. The DCNN will then learn from the labeled images and attempt to intuit which unlabeled images are also S-300 platforms. In this way, DCNN algorithms build robust architecture during multiple iterations of epoch training for the purpose of facilitating object detection in the real world. Now that this research paper has summarized the three methods of training DCNNs, it will transition to describing DCNN architecture.

Deep Convolutional Neural Network Architecture

DCNN architecture consists of two components (feature learning and classification) which support four distinct layers (convolution, activation, pooling and fully connected) that enable the algorithm's ability to detect objects. Refer to Figure 1 "DCNN Architecture" for a graphical depiction of DCNN architecture. We will move left to right beginning at the input image:

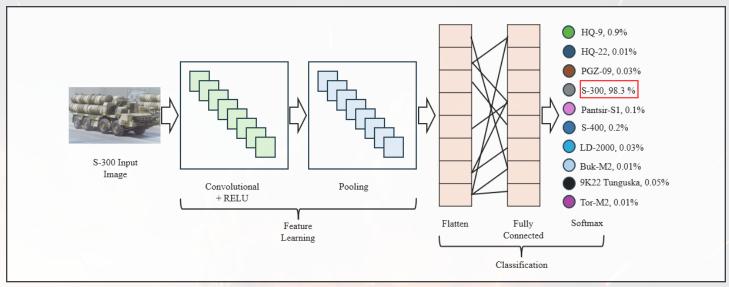


Figure 1. DCNN Architecture. Source: Created by Author

First, a DCNN will encounter an input image via still picture, video, electronic signature or some other measurable phenomena present in the real world. The first DCNN layer that will interact with the input image is the convolutional + rectified linear unit, or (ReLU) layer. This layer interacts with the input image by running a frame, sometimes referred to as a sliding window, over the input image in order to initially extract features from the input image. From this extraction, the DCNN begins building a quantitative feature map of the image that the ReLU layer can then exploit.

The ReLU layer acts as an activation function that assists the DCNN in sorting through the feature map.¹³ Figure 2 portrays how the ReLU layer numerically filters through input image features that it deems important or not important for object detection. This importance is defined by the prior epoch training that the DCNN has undergone. Positive values remain the same while negative values are automatically assigned a zero value. In this way, ReLU activation functions conserve computational power while simultaneously setting the conditions for the subsequent pooling layer.

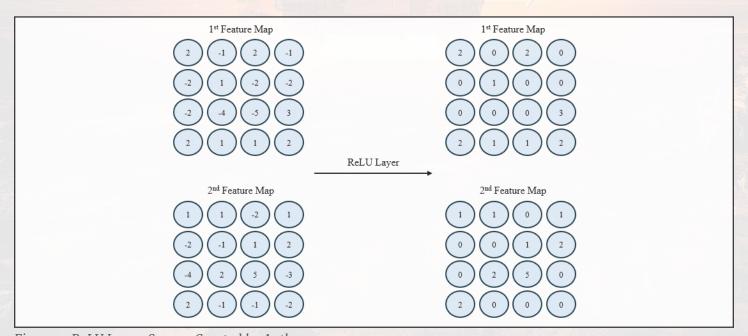


Figure 2. ReLU Layer. Source: Created by Author

The pooling layer interacts with the values assigned by ReLU in order to isolate important features identified by numerical value as well as systematically minimizing the "spatial size" of the original feature map. ¹⁴ In this way, pooling layers capitalize on the ReLU values in order to "reduce computational"

complexity" and begin extracting the primary features that constitute a known object. The result is a bounded box, as represented by Figure 3, that isolates an object from the background features of an input image. This bounded box represents the transition from the feature learning component to the classification component within DCNN architecture.

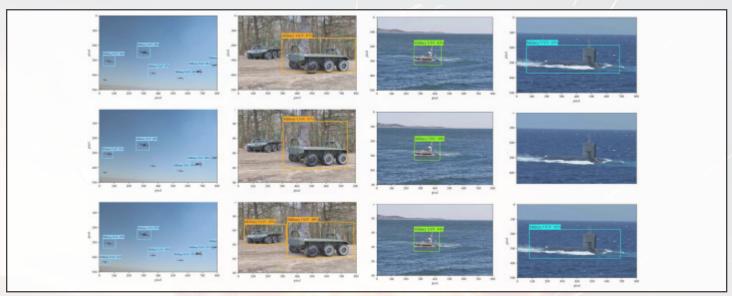


Figure 3. Bounded Box. Source: Liming Gao, Chao Li, Zhuo Wei, Xiangdong Han, Feng Dang and Xuemei Wei, "Military Unmanned Equipment Image Target Recognition Method based on Improved Deep Learning," Paper presented at the 2nd International Conference on Algorithms, Network, and Computer Technology, Wuhan, China, August 12-October 12, 2023, 5, DOI 10.1088/1742-6596/2732/1/012004.

Now that the DCNN has successfully isolated an object within its feature map, it can begin classifying the object within the bounded box. Classification is conducted in three steps: (1) flatten, (2) fully connected and (3) softmax. When a DCNN conducts flattening, it is essentially combining all the data points from the previous pooling layers which can amount to hundreds, thousands, hundreds of thousands or even millions of individual pooled feature maps. Consequently, the flattening phase of classification combines all the individual pooled feature maps into one long string of data which the neural network can begin processing. For example, Avijeet Biswal states that "flattening is used to convert all the resultant 2-Dimensional arrays from pooled feature maps into a single long continuous linear vector." Consequently, this linear vector contains the numerical code that the machine will use to ultimately classify an image in the subsequent fully connected layer.

The linear vector of data is now prepared to move through the fully connected layer. The fully connected layer receives the linear vector of data and processes it through interconnected neurons which remain linked to every previous and subsequent DCNN layer. ¹⁶ In other words, think of the fully connected layer as the layer which is doing the majority of the sensemaking because the fully connected layer is applying the linear vector of data to the recollection knowledge it has attained through previous epoch training. In this way, the linkage between the flatten and the fully connected stages produces a "class score" for the object within a bounded box, which represents the likelihood that an isolated image matches with a previously trained object. ¹⁷

From the class score, the final softmax application derives a probability statistic for each class in which the DCNN has been trained. If the probability statistic associated with any classification rating reaches the minimum threshold for positive identification, then the machine will assume that the bounded object it is looking at is that type of classification. Let's turn now to the current state of target acquisition DCNNs in order to determine their current state of readiness in the contemporary military domain.

Previously Constructed Target Acquisition DCNNs

Perhaps the most obvious use of DCNN disruptive technology for the military sphere is its ability to rapidly acquire and classify potential targets from video input. For example, B. Janakiramaiah et al. recently conducted a study that illustrates the efficacy of DCNNs for military target acquisition. ¹⁸ Their multi-level CapsNet DCNN trained on 600 images of armored cars, multi-barrel rocket systems, tanks, fighter planes and helicopter gunships as well as 500 images of non-military general objects for a grand total of 3,500 input images. Following epoch training, their DCNN achieved a 96.54% accurate target identification rate. Consequently, these researchers argue that their DCNNs provide a viable option for automatic target identification in contemporary armed conflict.

Similarly, building upon the previous ideas offered by Janakiramaiah et al., Guozhao Zeng et al., operating out of the Chinese National University of Defense Technology, successfully realized a portable DCNN for target acquisition. Their 15-layered-DCNN trained on six military objects and achieved an average 75% accurate target identification rate. The researchers conclude that their DCNN offers the Chinese military a viable option for military object detection.

And lastly, Anishi and Uma Gupta built a DCNN that is capable of differentiating between tanks, rifles, people, cars and trucks using images derived from regular daylight hours as well as images captured with night vision.²⁰ Their model consisted of 58 convolutional layers with five pool layers and demonstrated the need for high computing power to manage increasingly larger datasets. Therefore, this study, as well as the two previously mentioned studies, illustrates how DCNNs may contribute to the future of military target acquisition.

Bringing It All Together

Now that this article has offered a general sense of what DCNNs are, how they are trained, how they operate and how they are currently performing in the contemporary military domain, it will transition to describe why it matters for the fires warfighting function. Take, for example, John Boy's OODA loop (observe. orient, decide, act) concept and apply it to how human staffs at the division, corps or Army level move through the targeting process.

In his book A Discourse on Winning and Losing, Boyd describes how the OODA loop, colloquially known as "The Big Squeeze," facilitates a competitive thinking process capable of adapting to "an unfolding, evolving reality that is uncertain, ever changing and unpredictable." Figure 4 illustrates how one moves through the observation, orientation, decision and action stages of Boyd's concept. For years, human agents in all occupations have relied on this simple yet effective framework to update actions based on what is being observed in a chaotic environment.

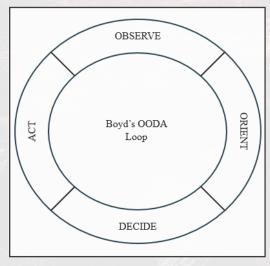
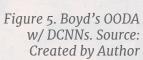
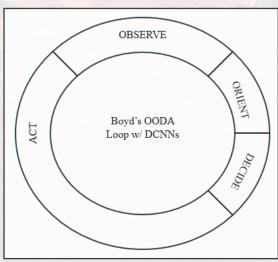


Figure 4. Standard Boyd OODA. Source: Created by Author





Consider how an Army staff facilitates the targeting process. Intelligence sections sift through massive amounts of observed data in the form of reports. Gun, rocket, rotary and fixed wing combat power is oriented based on placement of the coordinated fire line (CFL) and fire support coordination line (FSCL). The staff then works in conjunction with the commander to decide what/ where/when targets must be destroyed. Finally, action occurs. This is a continuous process reliant on copious amounts of human effort over a repetitive daily cycle. Certainly, this process has worked in the informatized age, absent of emerging artificial intelligence concepts. However, in the age of "intelligentized () warfare," as conceptualized by the Peoples Liberation Army, will our old mode of Army targeting be enough?²²

I would suggest the Fires community consider how targeting DCNNs can dramatically intensify the speed and responsiveness at which the deep fight is prosecuted. With the inclusion of future target acquisition DCNNs, Boyd's loop begins to look more similar to how it's depicted in Figure 5. The orientation and decision steps become much tighter leaving more room for spontaneous action at lower levels of command.

Target acquisition DCNNs can be constructed for intelligence, surveillance and reconnaissance (ISR) platforms which would then flood the area of operations with machines that can identify size, type and activity of enemy combat power. This data could then be fed into an umbrella command and control DCNN whose softmax layer is cognizant of friendly firing unit locations/readiness levels and is therefore positioned to intuit which firing unit is best suited to engage the targets fed to it by the subordinate target acquisition DCNNs.

Think of this umbrella DCNN as the AI sensor-to-shooter tool called for by The U.S. Army in Multi-Domain Operations 2028 document. It asserts "the key to converging capabilities across all domains, the EMS [electromagnetic spectrum], and the information environment is high-volume analytical capability and sensor-to-shooter links enabled by artificial intelligence, which complicates enemy deception and obscuration through automatic cross-cueing and target recognition." In this way, target acquisition and decision support DCNNs offer a viable option for achieving intelligentized warfare. Now, for the sake of visualization, consider the below theoretical scenario portraying how DCNNs may assist a future joint task force (JTF) while operating in the South China Sea (SCS).

In the near future, the Joint All Domain Command and Control (JADC2) architecture will be empowered by DCNNs which are capable of fusing massive amounts of ISR data with available Army, Navy, Airforce, Cyber and Space fire support systems. Each component command of a JTF maintains and employs its own proprietary DCNN which has been trained to recognize and prioritize adversarial weapon systems operating in a bounded geographic area. These individual component DCNNs are actually portions of an umbrella DCNN which the JTF commander and his staff exploit during the joint targeting cycle to facilitate rapid target acquisition and prosecution.

The umbrella DCNN is instantly cognizant of both the potential targets which are being fed to it by the subordinate component DCNNs as well as the positions, ammunition allocations and readiness of all the JTF commander's fire support systems within the assigned combatant command. Consequently, the DCNN is able to rapidly suggest which targets should be actioned by which friendly fire support system—regardless of component—in order to achieve the most desired effect.

The task that the JTF currently finds itself in is perhaps best described by General Charles Flynn and Lieutenant Colonel Tim Devine in which new Army long range precision missiles such as the Precision Strike, Strategic Mid-Range and Long-Range Hypersonic missile augment the Navy and Airforce's ability to penetrate and dis-integrate adversarial A2/AD bubbles within the SCS.²⁴ Flynn and Devine portray further how Army fires assets can be positioned "on key terrain inside the first island chain [in order to hold] the adversary's critical capabilities at risk via cross-domain strike."²⁵ Therefore, the JTF commander may position fire support assets within the Philippines, Borneo, Taiwan or the Ryukyu islands.

Keeping this scenario in mind, the JTF Multi-Domain Task Force would be receiving its targeting data from the DCNNs of other air, maritime, cyber or space components which are being filtered through the umbrella DCNN maintained at the JTF commander level. Consequently, the joint targeting cycle would be rapidly enhanced by orders of magnitude because it is no longer tied to the 96-hour cycle which requires numerous layers of human interaction between multiple services who, at times, find themselves at odds with each other. Instead, the umbrella DCNN, which is cognizant of all known potential targets as well as the location, munition allocation and status of all friendly fire support assets, would simply select the best positioned asset to achieve the desired effect.

In this way, with the assistance of robust and tempered DCNNs, JADC2 could actually become a reality. Interservice rivalries, turf wars and personal grievances would fade into the background while performance optimization and speed of joint targeting would become preeminent. Of course, this is merely a theoretical conceptualization. However, it does illustrate how DCNNs can greatly increase the effectiveness of target acquisition within armed conflict.

Conclusion

Of course, this is only a theoretical scenario. However, it does illustrate how future conflict could become incredibly reliant on neural networks which can outperform staff processes rooted solely in stand-alone human cognition. This article strove to bring the idea of DCNNs that much closer to the U.S. Army Fires community. The use of DCNNs for targeting may be uncomfortable for some decision makers because DCNNs represent a disruptive technology which upends established modes of warfighting. However, although the nature of war never changes, its character surely does. DCNNs, automatic target acquisition, computer vision and the myriad other applications for military-centric AI will assuredly change the character of future warfare. I agree with Seneca's quote at the head of this article. We must embrace target acquisition DCNNs or suffer being dragged through inevitable defeat because of our unwillingness.

MAJ Joseph D. Schmid is a student of the Advanced Military Studies Program at Fort Leavenworth, Kansas. He has published previously across a wide variety of subjects to include Land Anti-Ship Missiles, Multi-Domain Operations and drone warfare. He holds graduate degrees in English, Military Studies, and Military Art and Science. He is currently pursuing a Master of Arts in Military Operations.

Endnotes

- 1. Robinson, John. "Enhancing Tactical Level Targeting with Artificial Intelligence." *Field Artillery Professional Bulletin* iss. 1 (2024): 10. https://www.dvidshub.net/publication/issues/69894.
- 2. MacDonald, Norine and George Howell. "Killing Me Softly." *Prism 8*, no. 3 (2019): 114. https://www.jstor.org/stable/10.2307/26864279
- 3. Bubeck, Sebastian, Varun Chandrasekaran, Ronen Eldan, Johannes Gehrke, Eric Horvitz, Ece Kamar, Peter Lee, Yin Tat Lee, Yuanzhi Li, Scott Lundberg, Harsha Nori, Hamid Palangi, Marco Tulio Ribeiro, and Yi Zhang. "Sparks of Artificial General Intelligence: Early experiments with GPT-4." Research Paper, Cornell University, April 2023. https://doi.org/10.48550/arXiv.2303.12712.
- 4. Aristotle. *On Memory and Reminiscence*. Translated by J. I. Beare. Cambridge, MA: MIT Internet Classics Archive, 1994. http://classics.mit.edu/Aristotle/memory.html.
- 5. Glasner, Ruth. "An Early Stage in the Evolution of Aristotle's Physics." *Studies in History and Philosophy of Science* 81 (June 2020): 24-31. https://doi.org/10.1016/j.shpsa.2020.01.010.
- 6. W. H. Burnham, "Memory, Historically and Experimentally Considered." *The American Journal of Psychology* 2, no. 1 (1888): 41, https://doi.org/10.2307/1411406.
 - 7. Aristotle, On Memory and Reminiscence, part I.
 - 8. Ibid., para. 19.

- 9. Gu, Xiaowei, Plamen Angelov, Ce Zhang, and Peter Atkinson. "A Semi-Supervised Deep Rule-Based Approach for Complex Satellite Sensor Image Analysis." *Institute of Electrical and Electronics Engineers* 44, no. 5 (May 2022): 2282. 10.1109/TPAMI.2020.3048268.
- 10. Suryawati, Endang, Hilman Pardede, Vicky Zilvan, Ade Ramdan, Dikdik Krisnandi, Ana Heryana, Sandra Yuwana, Budiarianto Suryo Kusumo, Andria Arisal, and Ahmad Afif Supianto. "Unsupervised Feature Learning-Based Encoder and Adversarial Networks." *Journal of Big Data 8*, no.118 (2021): 5. https://doi.org/10.1186/s40537-021-00508-9.
- 11. Shi, Weiwei, Yihong Gong, Chris Ding, Zhiheng Ma, Xiaoyu Tao, and Nanning Zheng. "Transductive Semi-Supervised Deep Learning Using Min-Max Features." Paper presented at the 2018 European Conference on Computer Vision, Munich, Germany, 8-14 September 2018. 313. https://doi.org/10.1007/978-3-030-01228-1_19.
- 12. Zhiqiang, W., and L. Jun. "A Review of Object Detection Based on Convolutional Neural Network," Paper presented at the 36th Chinese Control Conference, Dalian, China, July 26-28, 2017. 1104-1106. doi: 10.23919/ChiCC.2017.8029130.
- 13. Ren, Junsong, and Yi Wang. "Overview of Object Detection Algorithms Using Convolutional Neural Networks." *Journal of Computer and Communications* 10, no. 1 (January 2022): 117. DOI: 10.4236/jcc.2022.101006.
- 14. Yang, Rui. "Convolutional Neural Networks and Their Applications in NLP." *In Modern Approaches in Natural Language Processing*, Chapter 5. Munich: Ludwig-Maximilians Universität München, 2020. 5.1. https://slds-lmu.github.jo/seminar_nlp_ss20/index.html.
- 15. Biswal, Avijeet. "Convolutional Neural Network Tutorial." *Simplilearn*. Last modified November 7, 2023. https://www.simplilearn.com/tutorials/deep-learning-tutorial/convolutional-neural-network.
- 16. Saad Albawi, Tareq Mohammed, and Saad Al-Zawi. "Understanding of a Convolutional Neural Network." Paper presented at the 2017 International Conference on Engineering and Technology, Antalya, Türkiye, 2017. 5. DOI: 10.1109/ICEngTechnol.2017.8308186.
- 17. Zhang, Sanxing, Zhenhuan Ma, Gang Zhang, Tao Lei, Rui Zhang, and Yi Cui. "Semantic Image Segmentation with Deep Convolutional Neural Networks and Quick Shift." *Symmetry* 12, no. 427 (2020): 1413. https://www.semanticscholar.org/reader/27c4457c5e403d21a0ad00d707546eb3df5c5941.
- 18. Janakiramaiah, B., G. Kalyani, A. Karuna, L. V. Narasimha Prasad, and M. Krishna. "Military Object Detection in Defense Using Multi-Level Capsule Networks." *Soft Computing* 27 (2023): 1045-1059. https://doi.org/10.1007/s00500-021-05912-0.
- 19. Zeng, Guozhao, Rui Song, Xiao Hu, Yueyue Chen, and Xiaotian Zhou. "Applying Convolutional Neural Network for Military Object Detection on Embedded Platform." *In Computer Engineering and Technology*, edited by Weixia Xu, Liquan Xiao, Jinwen Li, and Zhenzhen Zhu, 131–141. Singapore: Springer, 2019. https://doi.org/10.1007/978-981-13-5919-4_13.
- 20. Gupta, A., and U. Gupta. "Military Surveillance with Deep Convolutional Neural Network." Paper presented at the 2018 International Conference on Electrical, Electronics, Communication, Computer, and Optimization Techniques, Msyuru, India, 2018. 1147-1152. DOI: 10.1109/ICEECCOT43722.2018.9001381.
- 21. Boyd, John. *A Discourse on Winning and Losing*. Maxwell AFB, AL: Air University Press, 2018, 383. https://www.airuniversity.af.edu/Portals/10/AUPress/Books/B_0151_Boyd_Discourse_Winning_Losing.PDF.
- 22. Kania, Elsa B. Chinese Military Innovation in Artificial Intelligence: Testimony before the U.S. China Economic and Security Review Commission Hearing on Trade, Technology, and Military-Civil Fusion. Washington, DC: Center for a New American Security, 2019. 1. https://www.jstor.org/stable/resrep28742.
- 23. US Army. *The U.S. Army in Multi-Domain Operations 2028*, Fort Belvoir, VA: Army Publishing Directorate, 2018. 38. https://adminpubs.tradoc.army.mil/pamphlets/TP525-3-1.pdf.
- 24. Flynn, Charles, and Tim Devine. "To Upgun Seapower in the Indo-Pacific, You Need an Army." Proceedings 150, no. 2 (February 2024): 40 https://www.usni.org/magazines/proceedings/2024/february/upgun-seapower-indo-pacific-you-need-army.
 - 25. Ibid., 40.