Environmental Cues in Thermal Images Impair Vehicle Identification Training: Simulated Thermal Imagery as a Potential Solution

John J. Graybeal
KINEX, Inc.
UNITED STATES OF AMERICA

Samuel S. Monfort
KINEX, Inc.
UNITED STATES OF AMERICA

Susan Harkrider
U.S. Army RDECOM CERDEC Night Vision and Electronic Sensors Directorate
10221 Burbeck Road, Fort Belvoir, VA 22060
UNITED STATES OF AMERICA

ABSTRACT

Thermal imagery provides distinct advantages on the battlefield for target detection and identification of combat vehicles. However, the use of thermal imagery to identify a vehicle—by the location of its engine, for example—requires extensive training. Because thermal imagery of "live" vehicles is relatively expensive to collect, training imagery tends to be restricted to a limited number of images per target. However, being trained on a limited number of images can incentivize learning irrelevant cues (such as image quality or distinctive background elements) that can only reliably discriminate between vehicles in the training set. Trainees can therefore pass assessments of learning without learning the visual cues that are important for identifying vehicles in the field. In contrast to imagery collected in the field, simulated thermal imagery is relatively inexpensive and can be created from many camera perspectives and with diverse background elements. We discuss our experiences with thermal imagery training in light of the cognitive processes that underlie object recognition. We suggest that modeling software could partially replace the need for images captured in the field, and that high-fidelity thermal models have the potential to vastly improve the effectiveness of vehicle identification training.

1.0 INTRODUCTION

Thermal sensors are a powerful tool for vehicle identification because they capture highly informative features of a vehicle, including thermal "hotspots," that are not visible to the naked eye. However, soldiers require extensive practice with thermal imagery to accurately identify vehicles in the field using thermal sensors. Unfortunately, only a limited supply of thermal imagery is usually available for use in training because it is both difficult and expensive to obtain. In this paper, we discuss the problems that limited supplies of thermal imagery pose to training and assessment. We first discuss some of our own experiences in target identification training and then review psychological and neuroscientific insights into how the brain learns to recognize vehicles. We also present data collected during a thermal imagery training session. We conclude that high-fidelity thermal modeling of vehicles may be a solution needed to fill gaps in our ability to train and assess vehicle identification skills.

1.1 NVESD Experiences Training Soldiers To Use Thermal Imagery

The U.S. Army's Night Vision and Electronic Sensors Directorate (NVESD) operates a human perception lab, which investigates research questions regarding human performance with emerging sensor technologies. Many of these perception experiments are conducted using images of thermal vehicles, so research participants often must be quickly trained to identify vehicles using thermal imagery.

Typically, images used for training were taken with a thermal sensor by NVESD employees. To collect thermal imagery, NVESD employees must travel to a military installation that possesses the relevant vehicle and capture images of the vehicle from a variety of ranges and aspects (i.e., viewing angles). This process becomes expensive quickly considering travel expenses, operating costs of the vehicles, and the large time commitment from NVESD personnel.

All of the images for a given vehicle are typically collected from the same sensor at the same location, so these images tend to share certain background features and have similar image qualities. For example, the images of an M1A1 may all have the same distinctive group of trees in the lower left of the image, while the images of a T-62 may be relatively barren with large hills in the background. If trainees learn to recognize these distinctive background elements, they no longer need to practice the thermal identification skills they learned in training. The likelihood of recognizing a background element increases when the limited supply of thermal imagery necessitates using the same set of images for both training and testing. The potential training problems posed by these unintentional, inherent background elements are the central focus of this paper.

Over the years, researchers at NVESD developed concerns regarding the extent to which soldiers were truly mastering vehicle identification with a limited pool of images. They noted that soldier performance on perception tests was usually lower than predicted by models. Soldiers' lower-than-expected performance led many researchers to wonder whether soldiers were learning to recognize critical thermal hotspots as well as they should. In particular, researchers suspected that soldiers were somehow learning the exact images rather than the cues necessary to truly identify vehicles in a variety of images and contexts. To provide insight into these issues, we next review relevant cognitive and neurobiological research on the processes underlying human attention.

1.2 How Does the Brain Learn to Recognize Vehicles During Training?

Object recognition is defined as the ability to assign labels to an object, and these labels can either be broad (e.g., vehicle vs. person) or highly specific (e.g., M1A1 vs. T-62) [1]. Although early cognitive models of object recognition have attempted to understand the fundamental processes underlying object recognition [2], scientists are still struggling to understand how the brain recognizes objects from multiple viewpoints in multiple settings [3].

Visual attention is controlled dynamically by two opposite mechanisms with distinct neurological underpinnings that work synergistically to optimize performance [4-6]. The first mechanism is often referred to as "bottom-up" processing, whereby attention is directed as a consequence of the qualities of the sensory information itself [5-6]. One example of this process is visual attention being involuntarily captured by a blinking alarm or by the movement of another car while driving. This attentional mechanism is thought to require very little effort, and research suggests that individuals can function in a bottom-up capacity for extended periods of time without growing tired [7-8].

Of course, visual attention can also be directed consciously, known as "top-down" processing [5-6]. In contrast to bottom-up processing, top-down processing takes deliberate effort and sustained effort on the part of an observer, which has a draining, tiring effect [7-8]. Therefore, although soldiers can be trained to direct their attention towards key thermal information, their attention will inevitably be drawn to other salient features of an image, such as the background. Especially as cognitive resources become depleted and the balance between top-down and bottom-up attention shifts, there is a risk that features in an image that capture attention automatically will increasingly prevail over those to which soldiers have been trained utilize.

Whether or not a soldier will learn to use background elements in thermal imagery identification tasks can be explained by two perceptual processes: *perceptual learning* and *attentional learning* [9]. Perceptual learning is the process of becoming more sensitive to subtle differences in the visual features of different categories (e.g., vehicles) [7], [10-12]. Although training should cause soldiers to become more sensitive to subtle differences in vehicle features, repeated exposure to the same stimuli may also cause them to become more sensitive to differences in background elements. Attentional learning involves learning to allocate attention to features of an image that are most helpful in categorization [13-18]. Although training teaches soldiers visual cues that reliably distinguish vehicles, there is a risk soldiers will learn to allocate their attention to background cues that also reliably discriminate vehicles. The involuntary tendency for soldiers to detect background elements coupled with their increasing sensitivity to them might cause soldiers to fixate on background elements as a means of distinguishing vehicles, especially when they are presented with images of difficult-to-identify targets but salient, easier-to-recognize backgrounds.

To determine whether soldiers participating in NVESD quarterly testing were primarily using thermal hotspots or other extraneous background elements to identify vehicles, we designed a simple experiment to test soldiers' knowledge of thermal hotspots. We administered a thermal hotspot test after soldiers had already passed the traditional vehicle identification training assessment. We also administered a simple survey to gain additional insight into the strategies soldiers reported using during training and assessment.

2.0 METHOD

2.1 Participants

Ten soldiers were assigned to our perception lab for five days with the task of completing vehicle identification training and, to the extent to which they consented, participating in perception tests.

2.2 Training

In order to teach soldiers how to recognize thermal images of vehicles, we used the Recognition of Combatants – Vehicles (ROC-V) software, which contained images of various combat vehicles from all angles.

For the purposes of this experiment (and other perception tests conducted throughout the week), participants were trained to identify 12 different vehicles: six tanks (M1A1, M60A3, M551, T-55, T-62, T-72), three armored personnel carriers (M2A2, M113, BMP-2) and three self-propelled guns (M109A5, 2S3, ZSU-23-4). Training was considered complete when the soldiers were able to pass a difficult assessment of their ability to recognize vehicles. After training for several hours, soldiers were allowed to attempt to pass the training assessment. Soldiers continued alternating freely between training and attempting to pass the training assessment until they were successful.

2.3 Assessments

2.3.1 Training Assessment

The training assessment consisted of 97 thermal images presented in a random order, which always contained one of the twelve vehicles soldiers had been trained to identify. The test displayed vehicles at a variety of aspects and ranges. Soldiers had effectively unlimited time to select the correct answer (1,000 seconds), and were required to achieve a 96% accuracy score to pass. Soldiers who failed to pass were required to re-take the assessment until they did. Many of the images were challenging and involved vehicles located at distant ranges. Thus, many of the visual cues that are helpful to vehicle identification, such as the number of wheels, were either missing or degraded.

After completing the test, soldiers received feedback regarding which images they correctly/incorrectly identified. The feedback included another opportunity to view each test image, and in the case of an incorrect response, a comparable image of the vehicle that the soldier incorrectly selected.

Due to the limited supply of thermal imagery available, the training assessment was created from the same images used for training and it always contained the same exact images regardless of how many times a soldier needed to take the test to pass it. All images for an individual vehicle were collected from only a few locations, resulting in unintentional background similarities. Because the 12 vehicles were imaged at a variety of locations, these extraneous cues were generally unique to each vehicle.

2.3.2 Thermal Hotspot Test

To test knowledge of thermal hotspots directly, participants took the thermal hotspot test shortly after passing the training assessment. The hotspot test assessed knowledge of the two primary thermal hotspots: engine location and exhaust location. The two-dimensional space of each vehicle was broken down into a three-by-three grid; soldiers were asked to place a single check mark in the three-by-three grid to indicate the location of each thermal hotspot (**Figure 1**).

M1 Abrams: ENGINE

	Left	Center	Right
Front			
Middle			
Rear			

M1 Abrams: EXHAUST

	Left	Center	Right
Front			
Middle			
Rear			

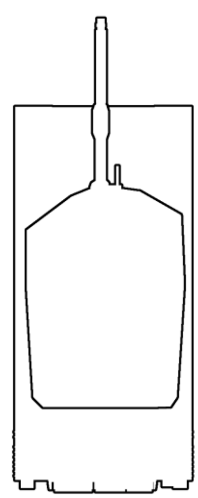

Figure 1: A sample question illustrating the format of the Thermal Hotspot Test. Participants had to place a checkmark in the correct grid location to indicate each hotspot. An outline of the vehicle was presented for reference and to facilitate mentally aligning the location labels with the vehicle's image.

Although most thermal hotspots fit nicely into discrete grid locations, there was ambiguity for several of the vehicles as to whether a hotspot fit only into a single grid location or if it could accurately be described by more than one grid location (i.e., some might have naturally described the hotspot on a vehicle as either "Right-Front" or "Right-Center"). In such cases, all potentially correct answers were scored as a correct response.

2.4 Debriefing Survey

To complement our objective learning assessments, we administered a brief exit survey to gather feedback about the soldiers' perceptions of the training assessment and on the cognitive strategies they used to pass it. In particular, we were interested in whether soldiers believed that the background elements were helpful to their identification efforts, and whether they ever intentionally studied the background in order to pass. The questions used a standard five-point Likert scale with anchors ranging from "Strongly Disagree" to "Strongly Agree."

3.0 RESULTS

3.1 Training Assessment

None of the soldiers passed the training assessment by the end of the first day of training, but all of them passed by the end of the second day. Most soldiers required at least three attempts to pass the training assessment and none passed it on the first attempt. The total amount of time it took each soldier to complete the training and pass the training assessment was approximately ten hours, although it varied by several hours between soldiers.

3.2 Thermal Hotspot Test

Accuracy scores on the thermal hotspot test were fairly low. The average score was a 54% (SD=24%), with scores ranging from 21% to 88%. Of the ten soldiers, 40% scored at a 33% or lower, suggesting that the group as a whole did not learn the critical thermal cues conveyed by the training.

3.1 Debriefing Survey

Participants indicated strong and consistent agreement (M=4.0, SD=1.4) with the following statement: "The picture backgrounds (for example: trees, hills, patterns in the ground) helped me identify vehicles on the [training assessment]." Critically, 80% of the participants agreed with this statement.

Participants were more neutral (M=2.8, SD=1.4) regarding the following statement: "I intentionally studied the background of images in the [training assessment] in order to pass." Nonetheless, 50% of the soldiers agreed with this statement, suggesting that background elements were at least salient enough to result in a deliberate learning strategy for half of the soldiers.

4.0 DISCUSSION

The results of the current study indicate that the limited pool of thermal imagery used for training and assessment under our current protocol was causing critical problems. Despite being able to accurately recognize and correctly identify vehicles on the training assessment, participants did not have sufficient knowledge of the thermal hotspot cues, which should have played a large role in their ability to correctly identify the vehicles at long range. In short, the suspicions of many NVESD researchers were confirmed: soldiers were learning to pass the training assessment without mastering the intended training content and they were accomplishing this in part by using the image backgrounds.

While repeatedly viewing thermal images, soldiers will begin to learn the image backgrounds for two reasons: first, as discussed in the introduction, attention is not only controlled consciously—“bottom-up” processing means that soldiers’ attention will naturally be drawn to salient and interesting features in an image, including the image background. Second, researchers have been able to identify several brain regions within the brain that specialize in recognition of scenes and locations [19-21]. These regions process visual information automatically and are always active, even without conscious effort [22]. The primacy of scene processing helps to explain why most of the soldiers in our experiment indicated that they gained an advantage on the training assessment by using the image background, although only half of them reported intentionally studying it.

It is also important to consider how perceptual learning and attentional learning contributed to our results. Perceptual learning, the process of becoming more sensitive to visual features that distinguish different vehicles, is not limited to the thermal hotspots and other features soldiers are supposed to be focused on—soldiers also became adept at detecting subtle differences in the backgrounds of the images. Eye-tracking studies, and other methods of measuring gaze location, demonstrate that humans learn to spend more time looking at the features of images that are the most informative for classifying objects during a specific task (attentional learning) [13-14], [16], [23-24]. In most situations, it is advantageous for people to begin to focus on the features of an image that are most informative. However, in our training assessment, participants were required to identify thermal signatures at extreme ranges where most of the visual information from the vehicle was degraded; small “white blobs,” on the horizon (thermal signatures at great distances) provided little information regarding what type of vehicle was present in the image. This lack of visual information regarding the vehicle, combined with soldiers’ imperfect knowledge of the hotspots and the consistency in vehicle backgrounds, naturally resulted in a shift in attention to background imagery. Because all of the images for a given vehicle came from the same unique environment(s), the environment ironically began to offer more information about vehicle type (i.e., which vehicle was displayed) than the image of the vehicle itself.

In summary, what we know about how the brain conducts object recognition coupled with the results of our experiment clearly indicate that not only were soldiers using the background of images to artificially pass the training assessment, but that this shift in cognitive strategy was a consequence of the way that the training assessment was conducted. These findings articulate the importance of developing training assessments that control for the imagery background (and other cues) so that extraneous cues cannot provide reliable information that helps soldiers pass a target recognition test.

4.1 Modeling Solves Inherent Practical Limitations of Authentic Thermal Imagery

Unfortunately, assembling a set of images that appropriately control for extraneous cues would require a prohibitively expensive number of thermal images. Although it might seem practical to capture images of all of the vehicles at a single location, our experiments often require images of vehicles that are rarely available at the same location. Worse, even capturing the same vehicles in the same location can result in unfortunate cues in the background. For example, the vehicles must be driven and parked at different aspects in order to image the vehicle from different angles. Over time, the ground becomes progressively more disturbed and thermal tracks become progressively added to image backgrounds. Differences in weather conditions may also affect a captured image's background; in short, it takes a long time to capture such imagery and the background environment is never entirely stable.

Given that it is impractical to capture vast numbers of images with ideal environmental background characteristics for use in thermal training and assessment, the use of high-fidelity thermal models may represent an alternative. Artificial images generated from thermal models have the potential to dramatically improve the assessment of a soldier's ability to recognize combat vehicles and should play an integral role in training and assessment. In particular, high-fidelity thermal models have the potential to accurately convey visual features of vehicles necessary for identifying their thermal signature while adding sufficient variance or perfect constancy to the backgrounds of thermal images; either perfect constancy or sufficient variance can be used to prevent soldiers from learning extraneous cues, but neither are realistically obtainable during field collects. Thus, training and assessment using simulated thermal imagery could prevent soldiers from artificially passing training assessments if they have not mastered the appropriate visual features that allow vehicle identification.

4.2 Other Advantages of Digital Modeling for Training and Assessment

In addition to solving problems inherent to extraneous cues present in thermal imagery, thermal modeling has the potential to improve the comprehensiveness of training by allowing soldiers to see the full effects of variables (e.g., range) that are impossible to represent fully in a small set of images. For example, thermal images of vehicles are captured at discrete ranges and aspects. Thermal modeling, on the other hand, would allow soldiers to view the vehicles from any aspect and range. Likewise, modeling software could include other functionality to alter image difficulty, such as being able to continuously degrade image quality by adding blur and noise.

Other important dynamic processes are also poorly represented in discrete images. For example, thermal signatures change as a function of engine-run time. When a vehicle is turned on, it takes some time for the engine, exhaust, and surrounding thermal cues to heat-up. Likewise, drive sprockets, wheels and tracks may heat-up while the vehicle is driving but may remain relatively cool while stationary. Thermal signatures also vary as a function of ambient temperature, weather conditions, day/night cycle, and other environmental variables. A tank that has been sitting in the hot sun all day, for example, will have a different thermal signature than one that has been sitting in an overcast and frigid region. It is not practical to capture these changes in thermal signature across time from many viewing aspects at many ranges during a field collect, but high fidelity software could model these different situations effectively and would better prepare soldiers by giving them a more comprehensive expectation of what a given vehicle looks like.

Finally, it is important to note variants of the same vehicle exist. For example, armored personnel carriers may or may not have an anti-tank missile launcher while tanks may or may not have track skirts covering the wheels. Variant features have the potential to significantly change soldier identification strategies, particularly when soldiers are inexperienced and still learning how to identify vehicles. Again, including vehicle variants in training that uses simulated thermal images is much more feasible than collecting thermal images of vehicle variants during a field collect. Advanced thermal imaging models may even be able to provide a useful thermal signature approximation for vehicles that are entirely unavailable for a field collect, such as new enemy vehicles not yet encountered in combat.

A logical extension of these capabilities of digital modeling is highly individualized training: trainers could rapidly prepare training material for highly specific combat theaters. By integrating all of these environmental factors, soldiers could be quickly trained on what specific vehicles may look like in a specific region. Instead of (or in conjunction with) having participants review real images of a vehicle collected halfway around the world, soldiers could review simulated imagery of vehicles from an appropriate terrain, climate, etc. High fidelity modeling could eventually even support mission specific training tasks within a combat theater, as soldiers could review the expected enemy and friendly thermal signatures, given the projected environmental conditions, immediately prior to the mission.

Limited data exists on human performance with simulated thermal imagery and it will certainly depend on the fidelity of the model, but some previous research suggests that human performance with simulated images will be similar to performance with actual imagery. For example, NVESD has previously investigated the consequences of using modeled versus actual thermal imagery of vehicles [25]. This investigation showed that observer performance on the simulated imagery was roughly similar to the performance of using real imagery in both the field and the lab, albeit that identification accuracy was somewhat higher for simulated images, which they attributed to higher contrast and less clutter in modeled imagery. As future iterations of thermal models are developed, it is likely that observer performance with simulated imagery will better approximate human performance with real imagery.

4.3 Conclusion

Our experience training soldiers to identify vehicles using thermal imagery coupled with our review of object recognition processes suggests that there are practical limitations to using real imagery for training and assessment purposes. Our results showed that soldiers were able to pass a difficult test of vehicle identification while still lacking critical knowledge of thermal hotspots by learning the background of the test images. The prohibitive cost and practical issues of collecting thermal imagery can lead to image sets where there is little variance in the background of images for a particular vehicle, which soldiers may learn to use (either consciously or subconsciously) to pass training assessments. Thermal imagery modeling is poised to resolve these practical and economic constraints by preventing soldiers from learning extraneous cues as these irrelevant image elements can easily be sufficiently varied or held constant, all while incorporating other environmental variance to the images that cannot be captured in field collects (e.g., weather conditions, engine run-time, time spent absorbing sunlight). As modeling technology continues to improve, thermal models will become an even more powerful tool for training and assessing vehicle identification skills and their use should become more frequently employed and evaluated empirically.

REFERENCES

[1] DiCarlo, J. J., Zoccolan, D., & Rust, N. C. (2012). How does the brain solve visual object recognition? *Neuron*, *73*(3), 415-434.
[2] Biederman, I. (1987). Recognition-by-components: a theory of human image understanding. *Psychological Review*, *94*(2), 115.
[3] Karimi-Rouzbahani, H., Bagheri, N., & Ebrahimpour, R. (2017). Hard-wired feed-forward visual mechanisms of the brain compensate for affine variations in object recognition. *Neuroscience*, *349*, 48-63.
[4] Sarter, M., Givens, B., & Bruno, J. P. (2001). The cognitive neuroscience of sustained attention: Where top-down meets bottom-up. *Brain Research Reviews*, *35*(2), 146-160.

[5] Connor, C. E., Egeth, H. E., & Yantis, S. (2004). Visual attention: Bottom-up versus top-down. *Current Biology*, *14*(19), R850-R852.
[6] Buschman, T. J., & Miller, E. K. (2007). Top-down versus bottom-up control of attention in the prefrontal and posterior parietal cortices. *Science*, *315*(5820), 1860-1862.
[7] Shiffrin, R. M., & Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending and a general theory. *Psychological Review*, *84*(2), 127.
[8] Smit, A. S., Eling, P. A., & Coenen, A. M. (2004). Mental effort causes vigilance decrease due to resource depletion. *Acta Psychologica*, *115*(1), 35-42.
[9] Hammer, R., Sloutsky, V., & Grill-Spector, K. (2015). Feature saliency and feedback information interactively impact visual category learning. *Frontiers in Psychology*, *6*(74), 1-15.
[10] Goldstone, R. L. (1994). Influences of categorization on perceptual discrimination. *Journal of Experimental Psychology: General*, *123*(2), 178.
[11] Goldstone, R. L. (1998). Perceptual learning. *Annual Review of Psychology*, *49*(1), 585-612.
[12] Roelfsema, P. R., van Ooyen, A., & Watanabe, T. (2010). Perceptual learning rules based on reinforcers and attention. *Trends in Cognitive Sciences*, *14*(2), 64-71.
[13] Rehder, B., & Hoffman, A. B. (2005). Eyetracking and selective attention in category learning. *Cognitive Psychology*, *51*(1), 1-41.
[14] Rehder, B., & Hoffman, A. B. (2005). Thirty-something categorization results explained: selective attention, eyetracking, and models of category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*(5), 811.
[15] Sloutsky, V. M., & Fisher, A. V. (2008). Attentional learning and flexible induction: How mundane mechanisms give rise to smart behaviors. *Child Development*, *79*(3), 639-651.
[16] Blair, M. R., Watson, M. R., Walshe, R. C., & Maj, F. (2009). Extremely selective attention: eye-tracking studies of the dynamic allocation of attention to stimulus features in categorization. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*(5), 1196.
[17] Hoffman, A. B., & Rehder, B. (2010). The costs of supervised classification: The effect of learning task on conceptual flexibility. *Journal of Experimental Psychology: General*, *139*(2), 319.
[18] McColeman, C. M., Barnes, J. I., Chen, L., Meier, K. M., Walshe, R. C., & Blair, M. R. (2014). Learning-induced changes in attentional allocation during categorization: a sizable catalog of attention change as measured by eye movements. *PloS one*, *9*(1), e83302.
[19] Köhler, S., Crane, J., & Milner, B. (2002). Differential contributions of the parahippocampal place area and the anterior hippocampus to human memory for scenes. *Hippocampus*, *12*(6), 718-723.
[20] Epstein, R. A., & Higgins, J. S. (2006). Differential parahippocampal and retrosplenial involvement in three types of visual scene recognition. *Cerebral Cortex*, *17*(7), 1680-1693.
[21] Park, S., & Chun, M. M. (2009). Different roles of the parahippocampal place area (PPA) and retrosplenial cortex (RSC) in panoramic scene perception. *Neuroimage*, *47*(4), 1747-1756.
[22] Marcel, A. J. (1983). Conscious and unconscious perception: Experiments on visual masking and word recognition. *Cognitive Psychology*, *15*(2), 197-237.
[23] Baruch, O., Kimchi, R., & Goldsmith, M. (2014). Attention to distinguishing features in object recognition. *Visual Cognition*, *22*(9-10), 1184-1215.
[24] Kruschke, J. K., Kappenman, E. S., & Hetrick, W. P. (2005). Eye gaze and individual differences consistent with learned attention in associative blocking and highlighting. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*(5), 830.
[25] Hixson, J. G., Teaney, B. P., May, C., Maurer, T., Nelson, M. B., & Pham, J. R. (2017, May). Virtual DRI dataset development. In *SPIE Defense+Security* (pp. 101780Q-101780Q). International Society for Optics and Photonics.