ARL-TR-10285 ● Feb 2026

Development of a Pseudo-HDR Preprocessing Pipeline for the Improvement of LDR Saliency Algorithms

by Darius Jefferson II and Andre Harrison

NOTICES

Disclaimers

ARL-TR-10285 ● FEB 2026

Development of a Pseudo-HDR Preprocessing Pipeline for the Improvement of LDR Saliency Algorithms

Darius Jefferson II and Andre Harrison
DEVCOM Army Research Laboratory

REPORT DOCUMENTATION PAGE

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED START DATE	3. DATES COVERED END DATE
February 2026	Technical Report	October 2022	February 2024

4. TITLE AND SUBTITLE
Development of a Pseudo-HDR Preprocessing Pipeline for the Improvement of LDR Saliency Algorithms

5a. CONTRACT NUMBER	5b. GRANT NUMBER	5c. PROGRAM ELEMENT NUMBER
5d. PROJECT NUMBER	**5e. TASK NUMBER**	**5f. WORK UNIT NUMBER**

6. AUTHOR(S)
Darius Jefferson II and Andre Harrison

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)	8. PERFORMING ORGANIZATION REPORT NUMBER
DEVCOM Army Research Laboratory ATTN: FCDD-RLA-IB Adelphi, MD 20793	ARL-TR-10285

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
DISTRIBUTION STATEMENT A. Approved for public release: distribution unlimited.

13. SUPPLEMENTARY NOTES

14. ABSTRACT
High dynamic range (HDR) imagery has an expanded luminance range capable of more accurately capturing the full range of luminance in the real world. However, most computer vision (CV) models cannot easily perceive the very dark or very bright areas in real-world HDR environments. This is because these models are typically trained on low dynamic range (LDR) imagery, which precondition CV models to expect a much smaller ratio between the brightest and darkest pixel in an image. Datasets of annotated HDR imagery are essential to training HDR-capable CV models for operation in real-world environments.
Saliency models are a type of CV model that would benefit from being retrained on HDR imagery. HDR imagery provides more information to the model for more accurate predictions. From a U.S. Army perspective, these retrained models could be used on autonomous systems in the battlefield to better inform Soldiers about areas in their environment that contain potential threats. However, very few HDR saliency datasets exist and none are large enough to train a model from scratch. In this report, we describe the development of a pipeline that would convert LDR image datasets into pseudo-HDR to train an HDR-adapted saliency model.

15. SUBJECT TERMS
Military Information Sciences, high dynamic range, HDR, saliency, tone mapping, expansion operators, domain adaptation

16. SECURITY CLASSIFICATION OF: a. REPORT	b. ABSTRACT	C. THIS PAGE	17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UU	18

19a. NAME OF RESPONSIBLE PERSON	19b. PHONE NUMBER *(Include area code)*
Darius Jefferson II	(520) 691-5180

STANDARD FORM 298 (REV. 5/2020)
Prescribed by ANSI Std. Z39.18

Contents

List of Figures

1. Introduction

Autonomous systems may find it difficult to perceive environments that contain very bright or very dim areas. For example, entering a dark cave from a bright environment with the light emitted from the entrance as the only available light source. A person’s vision can adjust for these changes in luminance so that they are able to observe their surroundings fairly accurately. However, this is not the case with autonomous systems. The real world is visually represented as high dynamic range (HDR)—meaning that it contains a wide range of luminance. Humans are better able to perceive HDR content in real time. Autonomous systems often use computer vision (CV) algorithms for analyzing images of their surroundings captured with their onboard camera. Most CV algorithms are only able to interpret low dynamic range (LDR) images because that is all they have been trained on. However, the dynamic range of LDR images is limited, as they do not capture the full luminance range seen in real-world environments. In addition, it has been shown that many LDR-trained CV models perform poorly when evaluated on HDR imagery.[1]

It would be ideal for the U.S. Army to improve the perception capabilities of their autonomous systems on the battlefield. An initial step toward this would be to upgrade the cameras of these systems to HDR cameras. Then, by training the autonomous systems’ CV models on the HDR image data, they would be able to analyze HDR camera images captured from the battlefield. These images would provide the models with closer visual representations of the battlefield, which often contain areas with variable lighting. Extremely bright or dim areas captured from the HDR environment may appear as over-/underexposed in an LDR image due to its limited luminance range. Those areas would provide no information for an LDR-trained CV model to learn from, which is why Army autonomous systems should use CV models trained on HDR imagery instead. This would allow the autonomous systems to better inform Soldiers of their surroundings and increase their situational awareness.

A type of CV model that would be useful for Army autonomous systems—especially if they were enhanced with HDR image training—is a visual saliency model. Saliency models try to predict where in an image a human might or should look. Those predicted locations often contain important objects of interest, such as potential hazards, adversaries, or obstacles. Autonomous systems on the battlefield that use HDR-trained saliency models could help to better inform Soldiers of these potential dangers. However, an obstacle that presents itself is the severe lack of available HDR saliency datasets. We found only four publicly available HDR saliency datasets online and only half of those are image datasets. Two image datasets (HDR-Eye[2] and ETHyma[3]) only contain 57 images between them. Most CV models need to be trained from scratch using hundreds or thousands of images in order to perform well. In addition, we cannot use the few HDR images from those datasets to fine-tune the LDR-trained saliency models either, as their backbones were not designed to process higher-bit image input like HDR. To potentially circumvent these issues, we propose the development of a pipeline that converts existing LDR saliency and image classification datasets into pseudo-HDR image datasets and retrains saliency models and their backbones using these pseudo-HDR datasets. As part of this work, the models' backbones would also need to be modified so that they will be able to process HDR image input. From there, the pseudo-HDR image datasets, after being preprocessed into a readable format for the backbones, would be used to retrain the backbones and the models themselves (their fixation heads, specifically). After retraining the models, we plan to evaluate them using real HDR saliency image datasets. In this report, we describe the development of a pseudo-HDR preprocessing pipeline designed to accomplish this aforementioned task. Current development progress includes detailing the selection of multiple LDR and HDR image datasets, expansion operators (EOs), and saliency models, describing the randomization process used before image preprocessing, the LDR to pseudo-HDR image conversion, and the selection of the image preprocessing methods.

2. Pipeline Development

2.1 Overview

The components of the pseudo-HDR pipeline that have already been implemented can be seen in Figures 1 and 2. In Figure 1, the process behind retraining the saliency models' backbones is shown. An LDR image classification dataset—specifically, one that was used to train one of the saliency models' backbones—is selected. Then, it is converted into pseudo-HDR using each of the EOs (a.k.a., inverse tone mapping). Together, the output images from each EO form a mixed pseudo-HDR classification dataset. The pseudo-HDR images from that mixed dataset are then normalized using various preprocessing methods. This is so that the images can be read by the classification model's backbone. Afterward, the backbone of the classification model is retrained using the pseudo-HDR images.

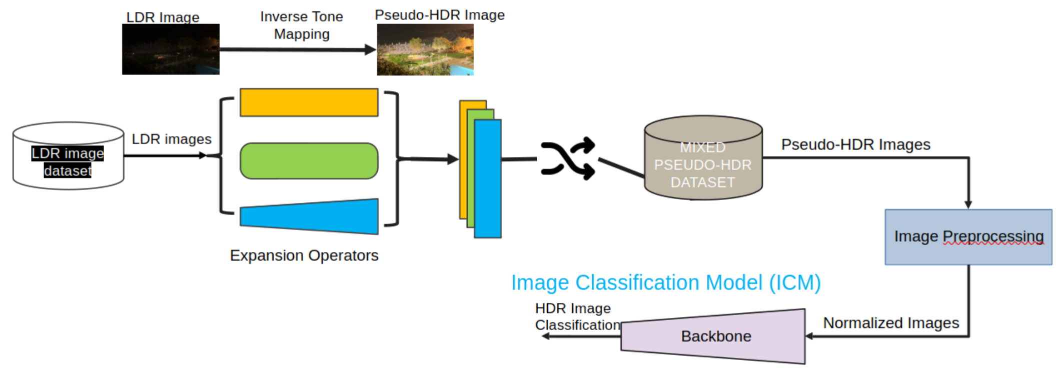

Figure 1. **A diagram displaying the first half of the pseudo-HDR preprocessing pipeline, which is the retraining of the saliency models' image classification backbone.**

The second half of the pipeline can be seen in Figure 2. In this half, the process behind retraining each of the saliency models is shown. The steps are similar to the steps from Figure 1. One difference is that LDR saliency datasets are being converted into pseudo-HDR instead of image classification datasets, because the saliency models are being retrained in this case. Another difference is that the saliency model backbone in Figure 2 is the same retrained pseudo-HDR image classification backbone from the classification model in Figure 1. The goal here is to retrain the saliency models' fixation heads using pseudo-HDR images converted from the LDR saliency datasets that the saliency models were originally trained on. Note that the steps in Figures 1 and 2 are repeated for each saliency model and their corresponding classification backbone model.

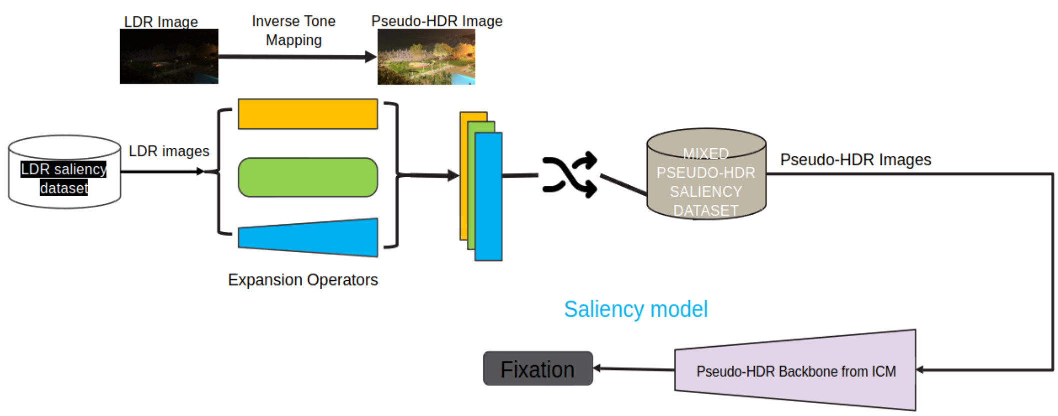

Figure 2. **A diagram displaying the second half of the pseudo-HDR preprocessing pipeline. This shows the retraining of the saliency models, which use the same pseudo-HDR retrained backbone from the image classification model shown in Figure 1.**

2.2 Current Progress

2.2.1 Visual Saliency Models

One of the first steps we completed when developing this pipeline was to select visual saliency models that we could later retrain using the pseudo-HDR images. When selecting each saliency model, we considered various factors. We prioritized models that were trained on LDR images, that ranked highly in several metrics of the Massachusetts Institute of Technology (MIT)/Tuebingen Saliency Benchmark,[4] and that could be easily implemented on our computers. For the four models we selected, we managed to successfully install, test, and confirm their functionality. The models we chose included Universal Image and Video Saliency (UNISAL),[5] Multi-Level Network (ML-Net),[6] Expandable Multi-Layer Network (EML-NET),[7] and TranSalNet.[8]

UNISAL unites static (image) and dynamic (video) saliency into a single model framework while also being computationally simple.[5] For its backbone, it uses MobileNetV2.[9] ML-Net uses a feature extracting convolutional neural network (CNN), a feature encoding network, and a prior learning network to predict saliency maps.[6] The backbone it uses is VGG-16.[10] EML-NET uses a multimodality system with prior knowledge to obtain a better saliency understanding.[7] This model uses a combination of two backbones. For the EML-NET model, the model's developers used NasNet-Large[11] and DenseNet-161.[12] However, on the model's GitHub repository, the developers used two differently trained versions of ResNet-50 instead.[13] TranSalNet unites the elements from transformers with CNNs for the purpose of acquiring long-range contextual visual information.[8] There are two versions of this model, each named for their model backbone. TranSalNet_Res uses ResNet-50 and TranSalNet_Dense uses DenseNet-161 (for TranSalNet we used TranSalNet_Res).

2.2.2 Datasets

The visual saliency models that were selected dictated which LDR saliency datasets would be used. We also identified the LDR classification datasets used for training each model's backbone. Almost all of the model's backbones (MobileNetV2, VGG-16, ResNet-50, NasNet-Large, and DenseNet-161) were trained on ImageNet.[14] The remaining backbones (i.e., alternate versions of ResNet-50 and DenseNet-161) were trained on the Places365[15] dataset.

For all saliency models selected, each was trained at minimum on a version of the Saliency in Context (SALICON)[16] dataset. Since UNISAL focused on using both images and videos, it was trained using more than just the SALICON dataset. In addition, UNISAL was trained on the Dynamic Human Fixation (DHF1K),[17] Hollywood-2,[18,19] and the University of Central Florida sports[19,20] datasets. Both EML-NET and TranSalNet were fine-tuned on the CAT2000[21] dataset. ML-Net, EML-NET, and TranSalNet were all fine-tuned on the MIT1003[22] dataset. All of these datasets, including the training datasets for the model backbones, will eventually be converted into pseudo-HDR to retrain the saliency models as part of our pipeline.

Additionally, we have selected two real HDR datasets that will eventually be used to evaluate the saliency models once they have been retrained: the HDR-Eye[2] and the ETHyma[3] datasets. All of the datasets that we have mentioned thus far—both LDR and HDR—have been downloaded.

2.2.3 EOs and Pseudo-HDR Image Conversion

To convert all of the images from the various LDR image datasets into pseudo-HDR, EOs were utilized. EOs (a.k.a., inverse tone mappers) expand the range of luminance values within LDR images. Doing so allows the expanded LDR images to emulate HDR imagery because real HDR images contain luminance values of a much wider range than LDR. In other words, these LDR images were converted into a pseudo-HDR format using the EOs.

For the pipeline, we decided to use multiple EOs to prevent any of the saliency models from overfitting on the output originating from a single EO. Each of the selected EOs is one of two types: deep neural networks (DNNs) or shallow. DNN EOs were used because they are often more accurate than shallow operators within their chosen domain, but not very generalizable. Shallow EOs were used because they are good at generalizing across multiple image domains; however, they are not as accurate as many DNNs. In the end, we chose four shallow and four DNN EOs. Most of the EOs that we selected were listed on the Moscow State University's HDR Video Reconstruction Benchmark[23] and could be implemented on our computers. We used the following DNN EOs: hdrcnn,[24] ExpandNet,[25] HDRUNet,[26] and TwostageHDR.[27] And, we used the following shallow EOs: Kovaleski–Oliveira,[28] Huo,[29] Kuo,[30] and Landis[31] methods. The specific versions of these methods that we used for the pipeline came from the HDR Toolbox,[32] which contained implementations of each function written in MATLAB.

The plan for converting the LDR datasets into pseudo-HDR involves dividing the images in each LDR dataset into nine sections. Eight of those sections would be converted by the eight EOs, while the ninth would remain as an LDR. The section placements for each image from the LDR datasets were determined using a random shuffle function. Randomly shuffling the images helps to reduce any biases that could occur from manually picking the sections. Using this conversion plan, we have currently converted the ImageNet and SALICON LDR datasets.

2.2.4 Image Preprocessing Methods

While investigating how to process HDR imagery using a CV algorithm, we examined how some tone mappers processed their HDR input before compressing them. We believe this may be an initial step needed to further convert the pseudo-HDR images into an interpretable format for retraining the backbones of LDR models. In the end, we decided to try to incorporate some of these methods for the pipeline by pursuing a similar process as seen in Mukherjee et al.[33] Currently, we have implemented six preprocessing methods in Python to be used for the pipeline. One is the linear conversion method, which normalizes a pseudo-HDR image such that its pixel values are 8-bit integers ranging from 0 to 255.

Another method is the gamma tone mapping linear conversion method, which is based on a method presented in Panetta et al.[34] In our implementation, a pseudo-HDR image has its pixel values normalized into 32-bit floating point representation, then tone mapped with the gamma value set to 2.1, and then normalized again using the aforementioned linear conversion method. The next method we implemented was floating point normalization, which normalizes a pseudo-HDR image such that its pixel values are 64-bit floating points ranging from 0 to 1. The fourth method that was implemented was the HDR pixel rescale method, which was based on a method in Mukherjee et al.[33] This method normalizes a pseudo-HDR image such that its pixel values are 64-bit floating points that range from 10^{-4} to 255. The fifth method was adaptive log compression, which originated from Vinker et al.[35] This method converts the red–green–blue (RGB) pixel values from a pseudo-HDR image into YUV color space values and then applies adaptive logarithmic compression to the luminance pixel values (i.e., the "Y" values). The sixth and final method was the illumination and reflectance filter extraction method, which is based on a method from Yang et al.[36] This method uses Laplacian pyramids to break down a pseudo-HDR image into its detail and base illumination components, the resulting outputs are effectively LDR versions of the pseudo-HDR image.

3. Conclusion

In this report, we described the development progress of a pseudo-HDR preprocessing pipeline. This includes the selection of several saliency models for retraining, LDR image datasets and EOs for pseudo-HDR conversion, and preprocessing methods. The next step planned for the pipeline is to finish converting all LDR training datasets into pseudo-HDR. After preprocessing those images, we will need to modify the backbones of the saliency models to accept HDR input for training. We would then be able to retrain the saliency models—both their backbones and fixation heads—using the converted and preprocessed pseudo-HDR images. Finally, we will evaluate these retrained models on real HDR saliency datasets (HDR-Eye and ETHyma) and compare their results to the original models evaluated on the same datasets.

In the future, we may consider modifying the pipeline based on an approach presented by Watanabe et al.[37] They propose an alternative approach for using HDR image content in LDR CV models. Rather than converting LDR images into pseudo-HDR, they convert HDR images into pseudo-LDR using tone mapping algorithms.[37] Specifically, they create multiple pseudo-LDR images by using different tone mapping parameters, sending each of them as input to an object detection model, and then merging each of their outputs using non-maximum suppression.[37] Their process mimics exposure bracketing—a camera technique that captures multiple LDR images of the same target at different exposures and merges them into one. In their case, however, they are doing the reverse.

4. References

1. Harrison A, Green M, Hung C, Raglin A. The consistency of visual search models on HDR and tone mapped images. 2019.
2. Nemoto H, Korshunov P, Hanhart P, Ebrahimi T. Visual attention in LDR and HDR images [presentation]. 9th International Workshop on Video Processing and Quality Metrics for Consumer Electronics (VPQM); 2015; Chandler, Arizona.
3. Narwaria M, Perreira Da Silva M, Le Callet P, Pepion R. Effect of tone mapping operators on visual attention deployment. In: Tescher AG, editor. Applications of Digital Image Processing XXXV; Proc. of SPIE 8499; 2012.
4. Kummerer M et al. MIT/Tuebingen saliency benchmark [home page]. [date unknown]. https://saliency.tuebingen.ai
5. Droste R, Jiao J, Noble JA. Unified image and video saliency modeling. Proceedings of the 16th European Conference on Computer Vision (ECCV); 2020.
6. Cornia et al. A deep multi-level network for saliency prediction. International Conference on Pattern Recognition (ICPR); 2016.
7. Jia S, Bruce NDB. EML-NET: an Expandable Multi-Layer NETwork for saliency prediction. Image and Vision Computing Vol. 95; 2020.
8. Lou J et al. TranSalNet: towards perceptually relevant visual saliency prediction. Neurocomputing. 2022.
9. Sandler M et al. MobileNetV2: inverted residuals and linear bottlenecks. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition; 2019; Salt Lake City, UT.
10. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. ICLR; 2015.
11. Zoph B, Vasudevan V, Shlens J, Le QV. Learning transferable architectures for scalable image recognition. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR); 2018.
12. Huang G, Liu Z, van der Maaten L, Weinberger KQ. Densely connected convolutional networks. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 2018.

13. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 2015.

14. Deng J et al. ImageNet: a large-scale hierarchical image database. IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 2009 June 20–25; Miami, FL.

15. Zhou B et al. Places: a 10 million image database for scene recognition. IEEE Trans Pattern Anal Mach Intell. 2018;40:6.

16. Jiang M, Huang S, Duan J, Zhao Q. SALICON: saliency in context. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 2015.

17. Wang W et al. Revisiting video saliency: a large-scale benchmark and a new model. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 2018.

18. Marszalek M, Laptev I, Schmid C. Actions in Context. IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 2009.

19. Mathe S, Sminchisescu C. Actions in the eye: dynamic gaze datasets and learnt saliency models for visual recognition. IEEE Trans Pattern Anal Mach Intell. 2015;37(7):1408–1424.

20. Rodriguez M, Ahmed J, Shah M. Action MACH a spatio-temporal maximum average correlation height filter for action recognition. 2008 IEEE Conference on Computer Vision and Pattern Recognition; 2008.

21. Borji A, Itti L. CAT2000: a large scale fixation dataset for boosting saliency research. IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 2015.

22. Judd T, Ehinger K, Durand F, Torralba A. Learning to predict where humans look. Proceedings of the IEEE International Conference on Computer Vision (ICCV); 2009.

23. MSU HDR Video Reconstruction Benchmark: HLG [homepage]. Moscow State University Computer Graphics and Multimedia Laboratory Video processing, compression and quality research group; last updated 2022 May 10. https://videoprocessing.ai/benchmarks/inverse-tone-mapping.html

24. Eilertsen G et al. HDR image reconstruction from a single exposure using deep CNNs. ACM Trans Graph. 2017;36(6).

25. Marnerides D, Bashford-Rogers T, Hatchett J, Debattista K. ExpandNet: a deep convolutional neural network for high dynamic range expansion from low dynamic range content. Computer Graphics Forum. 2018.

26. Chen X et al. HDRUNet: single image HDR reconstruction with denoising and dequantization. IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 2021.

27. Sharif SMA, Naqvi RA, Biswas M, Kim S. A two-stage deep network for high dynamic range image reconstruction. IEEE Conference on Computer Vision and Pattern Recognition Workshop (CVPRW); 2021.

28. Kovaleski R, Oliveira M. High-quality reverse tone mapping for a wide range of exposures. 27th SIBGRAPI Conference on Graphics, Patterns and Images; 2014; Rio de Janeiro, Brazil.

29. Huo Y, Yang F, Brost V. Dodging and burning inspired inverse tone mapping algorithm. J Comput Info Syst. 2013.

30. Kuo P, Tang C, Chien S. Content-adaptive inverse tone mapping. 2012 Visual Communications and Image Processing; 2012 Nov 27–30; San Diego, CA.

31. Landis H. Production-ready global illumination [course notes]. SIGGRAPH. 2002;16. p. 87–101.

32. Banterle F, Artusi A, Debattista K, Chalmers A. Advanced high dynamic range imaging. 2nd Ed. A K Peters/CRC Press; 2017.

33. Mukherjee R, Bessa M, Melo-Pinto P, Chalmers A. Object detection under challenging lighting conditions using high dynamic range imagery. IEEE Access. 2021;9.

34. Panetta K et al. TMO-Net: a parameter-free tone mapping operator using generative adversarial network, and performance benchmarking on large scale HDR dataset. IEEE Access; 2021;9.

35. Vinker Y, Huberman-Spiegelglas I, Fattal R. Unpaired learning for high dynamic range image tone mapping. Proceedings of the IEEE International Conference on Computer Vision (ICCV); 2021.

36. Yang J et al. Deep reformulated Laplacian tone mapping. arXiv. 2021. arXiv:2102.00348 [eess.IV].

37. Watanabe T, Kawakami R, Tanaka M, Okutomi M. Object detection framework using multiple tone mappings on high-dynamic-range images. IEEE International Conference on Image Processing (ICIP); 2024.

List of Symbols, Abbreviations, and Acronyms

CNN	Convolutional Neural Network
CV	Computer Vision
DHF1K	Dynamic Human Fixation
DNN	Deep Neural Network
EML-NET	Expandable Multi-Layer Network
EO	Expansion Operator
HDR	High Dynamic Range
LDR	Low Dynamic Range
MIT	Massachusetts Institute of Technology
ML-Net	Multi-Level Network
SALICON	Saliency in Context
UNISAL	Universal Image and Video Saliency

1 (PDF)	DEFENSE TECHNICAL INFORMATION CTR DTIC OCA
1 (PDF)	DEVCOM ARL FCDD RLB CI TECH LIB