

Comparison of algorithms for contrast enhancement based on triangle orientation discrimination assessments by convolutional neural networks

Daniel Wegner* and Stefan Keßler

Fraunhofer Institute of Optronics, System Technologies and Image Exploitation,
Ettlingen, Germany

Abstract. Within the last decades, a large number of techniques for contrast enhancement has been proposed. There are some comparisons of such algorithms for few images and figures of merit. However, many of these figures of merit cannot assess usability of altered image content for specific tasks, such as object recognition. In this work, the effect of contrast enhancement algorithms is evaluated by means of the triangle orientation discrimination (TOD), which is a current method for imager performance assessment. The conventional TOD approach requires observers to recognize equilateral triangles pointing in four different directions, whereas here convolutional neural network models are used for the classification task. These models are trained by artificial images with single triangles. Many methods for contrast enhancement highly depend on the content of the entire image. Therefore, the images are superimposed over natural backgrounds with varying standard deviations to provide different signal-to-background ratios. Then, these images are degraded by Gaussian blur and noise representing degradational camera effects and sensor noise. Different algorithms, such as the contrast-limited adaptive histogram equalization or local range modification, are applied. Then accuracies of the trained models on these images are compared for different contrast enhancement algorithms. Accuracy gains for low signal-to-background ratios and sufficiently large triangles are found, whereas impairments are found for high signal-to-background ratios and small triangles. A high generalization ability of our TOD model is found from the similar accuracies for several image databases used for backgrounds. Finally, implications of replacing triangles with real target signatures when using such advanced digital signal processing algorithms are discussed. The results are a step toward the assessment of those algorithms for generic target recognition. © The Authors. Published by SPIE under a Creative Commons Attribution 4.0 International License. Distribution or reproduction of this work in whole or in part requires full attribution of the original publication, including its DOI. [DOI: [10.1117/1.OE.62.4.048103](https://doi.org/10.1117/1.OE.62.4.048103)]

Keywords: advanced digital signal processing; convolutional neural networks; image contrast enhancement; imager assessment; method comparison; triangle orientation discrimination.

Paper 20221420G received Dec. 8, 2022; accepted for publication Mar. 24, 2023; published online Apr. 20, 2023.

1 Introduction

For remote sensing applications and reconnaissance, acquisition and operation of cameras in different spectral bands are required, and each has its own pros and cons. The best possible choice among devices for procurement is therefore dependent on the imager performance for the desired task, e.g., the detection, recognition, or identification (DRI) of distant targets with a background composed of vegetation, urban structures, and sky. Camera data can be acquired in field trials for characterization of single devices. However, these measurements are time consuming and expensive. Furthermore, the possession of the device is required. Therefore, modeling and image-based simulation of imagers are useful and important for the assessment of imagers. Such tools become even more important when scene-dependent advanced digital signal processing (ADSP) techniques are used in the device, for their impact on performance is difficult to predict. In this paper, the effect of contrast enhancement (CE) algorithms, which have so far mainly evaluated in terms of esthetic perception, is considered.

*Address all correspondence to Daniel Wegner, daniel.wegner@iosb.fraunhofer.de

Triangle orientation discrimination (TOD)¹ is a well-established image-based approach for the characterization of electro-optical system performance, especially for range performance assessment in remote sensing applications.² It models the DRI tasks for real targets by a simplified recognition task. The original idea was that an observer has to determine the orientation of an equilateral triangle directing in four directions (up, down, left, and right) shown on a display, which is fed by an imaging system. The ability to clearly discriminate the orientation is reduced depending on different types of degradation, e.g., optical diffraction blur and sensor noise.

However, due to the resurgence of machine learning, automatic target recognition³ is becoming increasingly important. The importance of machine vision applications also led to the emergence of scalable compression frameworks^{4,5} aimed at high lossy compression while simultaneously preserving image quality for machine and human vision. In contrast to human observers, the performance of these methods does not depend on properties of the display but merely on the digital output of the imaging system. A prominent and frequently used approach for machine vision is convolutional neural networks (CNN). Such CNN models also have been trained on artificial images of triangles to perform the TOD discrimination and validated on acquired camera data.⁶ Therefore, these models can be used for automated camera tests in the lab by means of scene projectors.⁷

In this paper, CNN models for TOD discrimination are trained and validated on degraded artificial images of single triangles superimposed over natural backgrounds from Open Images V6.⁸ In addition, the training data are processed by CE algorithms from Table 1 with equal probabilities. Then, a trained model is validated on images with varying error levels of background $\sigma_{\text{background}}$ and Gaussian noise σ_{noise} . Parts of this work have already been published elsewhere.³²

In Sec. 2, the considered CE algorithms, the model setup, and the training procedure are described. Section 3 shows the accuracies on validation images with varying background and noise levels. Accuracy differences between individual CE algorithms and identity are shown for varying values of the signal-to-background ratio $\text{SNR}_{\text{background}}$, signal-to-noise ratio $\text{SNR}_{\text{noise}}$, and triangle circumradius r . Finally, results are discussed in Sec. 4, which concludes the paper.

2 Methods

2.1 Considered Contrast Enhancement Algorithms

Several methods for CE have been proposed within the past decades. Various modifications to conventional global histogram equalization have been proposed to counteract mean brightness shifts^{11,12} leading to annoying artifacts and allowing smooth transitions to identity.^{25,26} Also learning-based methods^{33–37} as well as methods based on image decomposition^{38,39} have been proposed. The decomposition often relies on color information, making the methods inapplicable to single channel image data. However, the scope of this work is limited to easily implemented algorithms, given in Table 1, that operate on single channel image data.

2.2 Data Generation

For the training of models for TOD,¹ images of triangles are generated with varying contrasts, sizes, and four orientations, i.e., up, down, left, and right. Misalignment angles uniformly distributed in $[-15 \text{ deg}, 15 \text{ deg}]$ are added to the orientation angles of the triangles to make the models more robust to misalignment. Exceeding the maximum rotation angle 15 deg would lead to incorrect labeling because rotations of an equally sided triangle by 30 deg result in other labeled orientations due to the 120 deg rotational symmetry. This rotation is crucial when applying models on real camera data because some misalignment between the field of view of a camera and a target is unavoidable.

2.3 Background Overlay

Background images are extracted from OpenImages V6⁸ as the random square crops. RGB images are converted to floating-point grayscale images. Mean μ_{crop} and standard error σ_{crop}

Table 1 25 methods for CE.

Algorithm	Author	Year
AGC ⁹	Huang et al.	2013
AGCWD ⁹	Huang et al.	2013
BBHE ¹⁰	Kim et al.	1997
BPDHE ¹¹	Ibrahim et al.	2007
BPHEME ¹²	Wang et al.	2005
CLAHE ¹³	Zuiderveld et al.	1994
DCRGC ¹⁴	Wang et al.	2009
DCTIE ¹⁵	Tang et al.	2003
DSIHE ¹⁶	Wang et al.	1999
EHS ¹⁷	Coltuc et al.	2006
ESIHE ¹⁸	Tan et al.	2019
FHSABP ¹⁹	Wang et al.	2008
MMBEBHE ²⁰	Chen et al.	2003
MMSICHE ²¹	Singh et al.	2014
MPHEBP ²²	Wangsritong et al.	1998
NMHE ²³	Poddar et al.	2013
QDHE ²⁴	Ooi et al.	2010
RMSHE ²⁵	Chen et al.	2003
RSIHE ²⁶	Sim et al.	2007
RSWHED ²⁷	Kim et al.	2008
RSWHEM ²⁷	Kim et al.	2008
SUACE ²⁸	Bandara et al.	2017
LDRHE ²⁹	Bulut et al.	2022
LRM ³⁰	Fahnestock et al.	1983
WMSHE ³¹	Wu et al.	2010

are calculated within these crops. Then, gray levels $I_{\text{triangle+overlay}}(x, y)$ of an image with single triangle and background overlay are calculated as

$$I_{\text{triangle+overlay}}(x, y) = \begin{cases} c_{\text{constant}} + a_{\text{triangle}} & (x, y) \in \text{triangle} \\ c_{\text{constant}} + \sigma_{\text{set}} \cdot \frac{I_{\text{crop}}(x, y) - \mu_{\text{crop}}}{\tilde{\sigma}_{\text{crop}}} & \text{else} \end{cases} \quad (1)$$

$c_{\text{background}}$ is a constant gray level over the entire image. a_{triangle} is an offset value only added for pixels related to the triangle. In Eq. (1), the pixel values of the image crop $I_{\text{crop}}(x, y)$ are normalized by subtracting the mean μ_{crop} and dividing by the corrected error, and

$$\tilde{\sigma}_{\text{crop}} = \begin{cases} \sigma_{\text{crop}} & \sigma > \text{thres} \\ 1 & \text{else} \end{cases}, \quad (2)$$

with $\text{thres} = 10^{-4}$. Without this correction, Eq. (1) is not well-defined for uniform image sections because in this case the denominator would be $\sigma_{\text{crop}} = 0$. Then, the normalized gray levels related to the cropped natural background are scaled to have a specific standard deviation σ_{set} . Therefore, the signal-to-background ratio is expressed as

$$\text{SNR}_{\text{background}} [\text{dB}] = 10 \log_{10} \left(\frac{a_{\text{triangle}}}{\sigma_{\text{set}}} \right). \quad (3)$$

2.4 Degradations

Several image degradations representing typical camera effects are applied. Temporal noise is applied as uncorrelated additive Gaussian noise. Fixed pattern noise of a sensor is modeled as line- and column-based additive Gaussian noise. Linear motion blur on the triangle is applied to represent moving targets. Stabilization errors due to camera vibration are applied as linear motion blur and Gaussian blur on the triangle with a background overlay. Blur due to optical diffraction by circular apertures is applied by filters

$$g(x, y) = (2J_1(r/s)/(r/s))^2, \quad (4)$$

with $r = \sqrt{x^2 + y^2}$. These filters represent Airy disks⁴⁰ as the diffraction patterns of a circular aperture. $J_1(x)$ is the Bessel function of the first kind and first order. To provide optical diffraction blur for varying values of aperture diameter D , detector pixel pitch p , wavelength λ , and focal length f , a dimensionless scaling factor s is introduced as

$$s = \frac{\lambda f}{\pi D p}. \quad (5)$$

A variety of physical parameters (λ , f , D , and p) can be realized by random sampling of s from a uniform distribution in $[0.1, s_{\text{max}}]$. $s_{\text{max}} = 10$ is chosen to limit the proportion of severely degraded images, which aggravate the model training due to little possible accuracy gains compared with statistical fluctuations. Optical diffraction blur is applied by spatial filtering with random 2D kernels of width and height $K = 6s_{\text{max}}$. The kernel size K is limited due to lack of information beyond the borders of images of finite size. The scaling factor $s \leq s_{\text{max}}$ is therefore also limited because higher values lead to radial kernel profiles biased by clipping effects due to the limited kernel size. To reduce aliasing due to the oscillatory form of the Airy disk [Eq. (4)], larger kernels f_{ij} of $f_{\text{os}}K \times f_{\text{os}}K$ pixels are generated with an oversampling factor $f_{\text{os}} = 8$. These extended kernels are downsampled by average pooling with this oversampling factor to give $K \times K$ kernels g_{ij} . Normalized filter kernels are then formed as

$$\tilde{g}_{ij} = \frac{g_{ij}}{\sum_{i'}^K \sum_{j'}^K g_{i'j'}}. \quad (6)$$

Several aliasing effects may occur due to small detector fill factors $ff < 1$ (the ratio between the detector dimension and pitch) or different shapes of detector footprints, e.g., rhombic or circular. These effects can be realized by masking extended kernels of Airy disks f_{ij} with the detector profile before average pooling. However, this option is not used in this work due to the rare use of nonsquare detectors, low signal-to-noise ratio for low fill factors,⁴¹ and faster image generation.

2.5 Contrast Enhancement

The algorithms in Table 1 are applied on 50% of the degraded image with equal probabilities. Due to the complex and divergent control flow of some algorithms, these methods are implemented and calculated by a separate application on the CPU. In Fig. 1, pristine training examples, as well as degraded and ADSP processed ones, are shown. These training examples are generated online during training to have a practically infinite amount of training data immune to

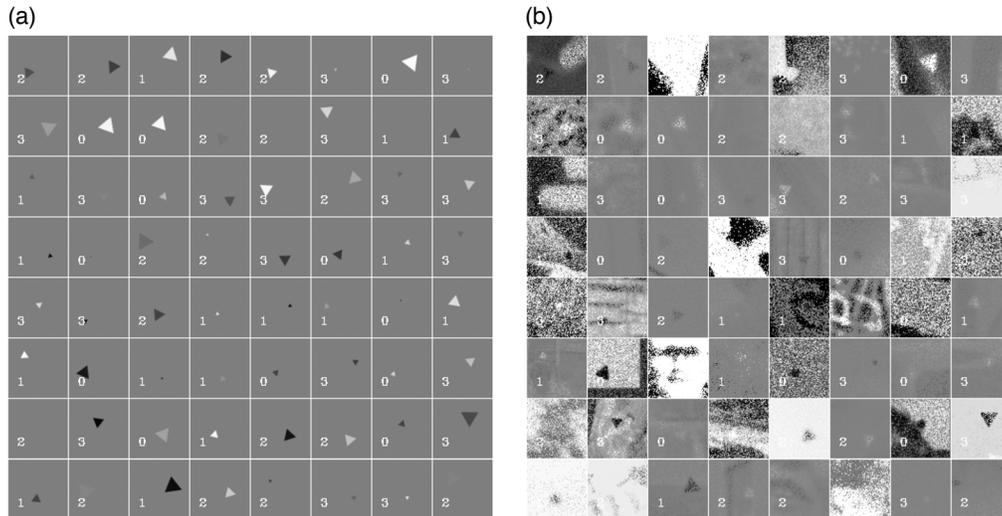


Fig. 1 Training examples with associated labels: left (0), up (1), right (2), and down (3). (a) Pristine training examples, (b) training examples with natural background, degradations, and CE.

overfitting. However, the source of background images and the number of possible crops is large but finite.

2.6 Model Setup

A conventional CNN architecture shown in Fig. 2 is used for TOD classification on images of dimensions $2^n \times 2^n$. To facilitate the model training, the input image is normalized by linear shifting and scaling of pixel values to have a mean of 0 and a standard deviation of 1 over spatial dimensions. Uniform input images with a standard deviation of 0 are not scaled. Then, the normalized image is downsampled by a chain of building blocks until the spatial dimensions are reduced to 2. Each building block consists of two 2D convolutional layers with 3×3 kernels and rectified linear unit activations (ReLU) and a subsequent 2×2 max pooling layer. Hence, downsampling by a factor 2 is applied per block. The spatial dimensions are reduced, and the number of feature maps, given as

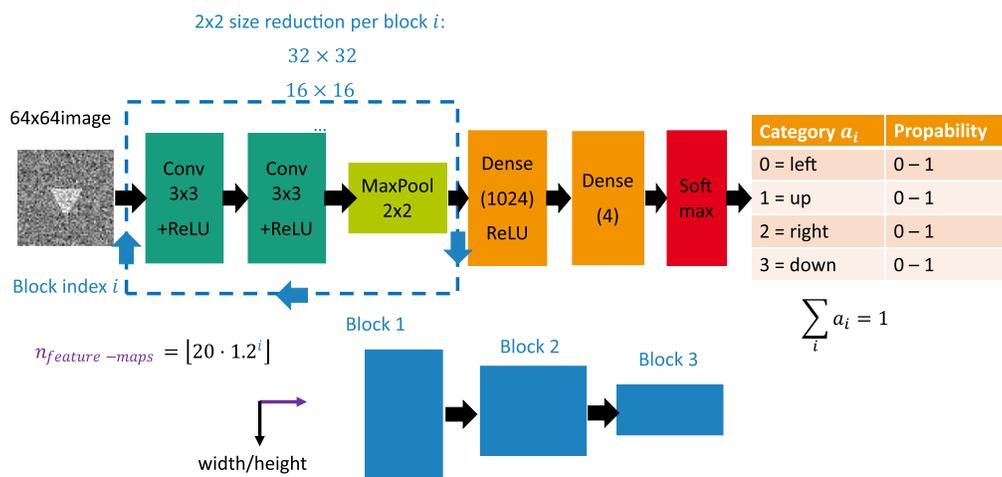


Fig. 2 CNN model architecture for TOD classification of 64×64 grayscale images. A 64×64 image is fed into a chain of blocks. Each block consisting of two 2D convolutional layers with ReLU activation and subsequent max pooling layer downsamples the spatial dimensions while increasing the number of feature maps. The last block is followed by two dense layers and a final softmax layer to provide probabilities for all directions (left, up, right, and down).

$$n_{\text{feature-maps}} = \lfloor N_0 \cdot g^i \rfloor, \quad (7)$$

increases for each block $i \geq 0$. $\lfloor \bullet \rfloor$ is the floor function yielding the largest integer that is smaller than the argument. Two subsequent dense layers and a final softmax layer provide the probabilities for the four orientations. A default configuration with the initial number of filters $N_0 = 20$, growth factor $g = 1.2$, and first dense layer size $L = 1024$ is arbitrarily chosen.

2.7 Model Training

The models are trained and evaluated by Python 3.9/Tensorflow 2.8. For optimization, ADAM⁴² with a learning rate $\eta = 0.001$ is used. Weights are initialized by He normal initialization.⁴³ Models are trained for $N = 1000$ iterations. Despite slower training, techniques for acceleration of training, such as batch normalization,⁴⁴ weight normalization,⁴⁵ or adaptive gradient clipping,⁴⁶ were deliberately omitted to achieve smaller models with faster inference, which are compatible for running on edge TPUs.⁴⁷ The loss function is cross entropy.

In each iteration, new sample images of triangles with background overlays are generated on the fly during the training for data augmentation. Triangles have a random size, position, and arbitrary orientation angle in $[0 \text{ deg}, 360 \text{ deg}]$. In addition, these images are impaired by the prescribed degradations. 50% of the training images were enhanced with one of the 25 methods for CE with equal probabilities. Corresponding disjoint sets of background images are randomly chosen from the respective partition of the OpenImages V6 database. For evaluation of model performance over the degradation parameters, i.e., $\text{SNR}_{\text{background}}$, $\text{SNR}_{\text{noise}}$, and triangle size, images are generated in the same way as background images chosen from the test subset of the database.

The question arises how large the percentage of degraded and ADSP processed images in the training data should be to obtain acceptable accuracies on validation sets of pristine and degraded images. 64×64 models with different percentages of degraded and processed images in the training data were trained and evaluated on different kinds of validation data. The respective accuracies on the validation data are shown in Fig. 3. It can be observed that models trained

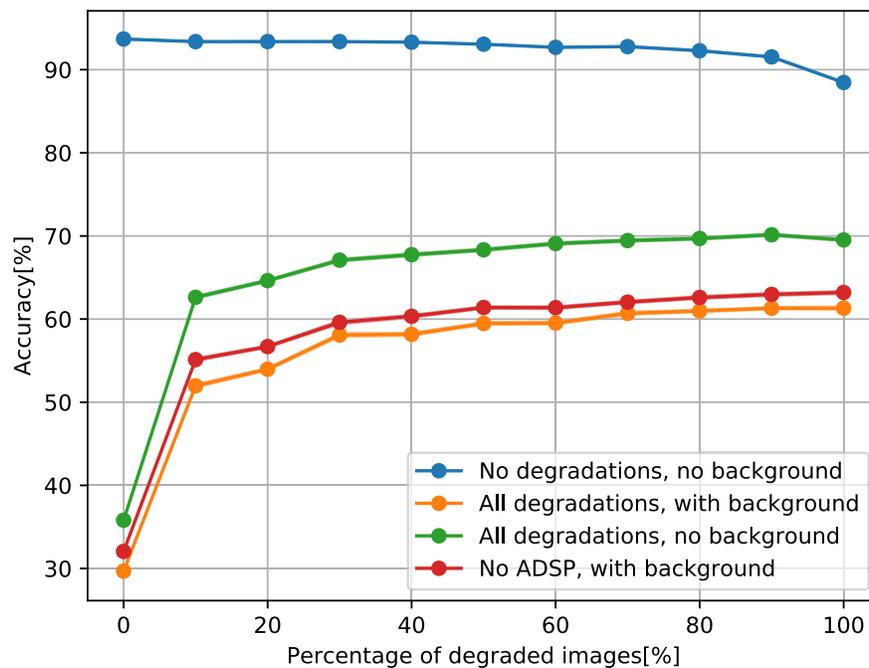


Fig. 3 Dependency of the validation accuracy on the composition of training sets. Horizontal: percentage of degraded images in the training data. Vertical: accuracies on different validation sets, each with $N = N_{\text{orientations}} \cdot N_{\text{backgrounds}} \cdot N_{\text{samples}} = 400,000$ images, with $N_{\text{backgrounds}} = 1000$, $N_{\text{orientations}} = 4$, $N_{\text{samples}} = 100$. “All degradations” include images enhanced with one of the algorithms in Table 1.

only with pristine images perform very bad on degraded and ADSP processed images with natural backgrounds. A slight increase of the percentage significantly raises the validation accuracies on degraded imagery. On the other hand, models trained with a high percentage of degraded images still perform very well on pristine images. Therefore, to make the best use of the model capacity for the image degradations and ADSP methods, all models mentioned below are trained with 100% degraded images, whereas ADSP is applied with 50% probability.

3 Results

3.1 Dependency of Accuracy on Background Variance

A trained 64×64 model is validated on images of a fixed target, a centered triangle with a circumradius of $r = 10$ pixel. The triangle circumradius r is converted to the often-used square root area $S = \sqrt{A}$ using the Pythagorean theorem:

$$S[\text{pixel}] = \sqrt{\frac{3\sqrt{3}}{4}r[\text{pixel}]} \approx 1.14r[\text{pixel}]. \tag{8}$$

1000 random crops of different background images from OpenImages V6⁸ were used. The background variance $\sigma_{\text{background}}$ of gray levels was varied to have different $\text{SNR}_{\text{background}} \in [0, 1, 2, 3, 4, 5, 10, 20]$ dB. White Gaussian noise was added to have $\text{SNR}_{\text{noise}} \in [0, 5, 10]$ dB, where

$$\text{SNR}_{\text{noise}} [\text{dB}] = 10 \log_{10} \left(\frac{a_{\text{triangle}}}{\sigma_{\text{noise}}} \right), \tag{9}$$

and a_{triangle} is about one-sixth of the dynamic range. In Fig. 4, accuracies over $\text{SNR}_{\text{background}}$ for different $\text{SNR}_{\text{noise}}$ and corresponding example images for the lowest $\text{SNR}_{\text{noise}} = 0$ dB are shown.

Obviously, accuracies are about 100% for $\text{SNR}_{\text{background}} \geq 5$ dB and $\text{SNR}_{\text{noise}} \geq 5$ dB. The accuracies drop monotonically with decreasing $\text{SNR}_{\text{background}}$. A similar behavior can be observed for varying triangle sizes, as shown in Fig. 5. As expected, the accuracies also drop for decreasing triangle sizes. Further variation of the relative triangle position in the subpixel range shows high fluctuations of accuracies for low triangle circumradius $r = 1$ pixel. This finding is consistent with the problem of recognition near the resolution limit mentioned before.¹

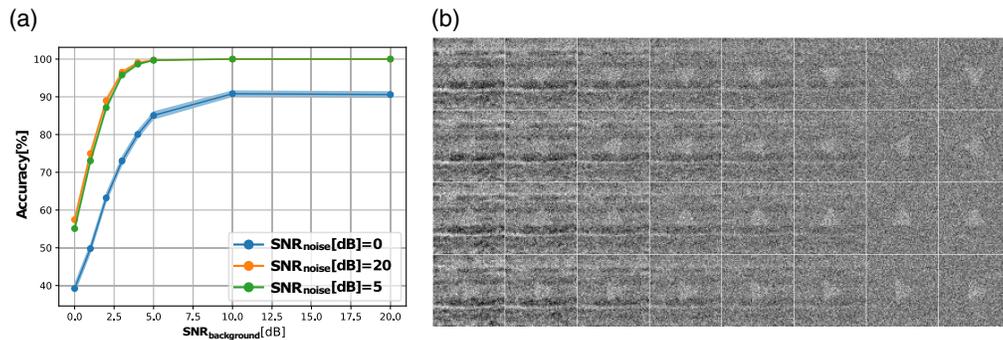


Fig. 4 (a) Accuracy over $\text{SNR}_{\text{background}}$ (dB): fixed triangle circumradius r (pixel) = 10 and varying $\text{SNR}_{\text{noise}}$ (dB). (b) Example validation images: fixed $\text{SNR}_{\text{noise}} = 0$ dB, four orientations and eight levels of $\text{SNR}_{\text{background}} \in [0, 1, 2, 3, 4, 5, 10, 20]$ dB.

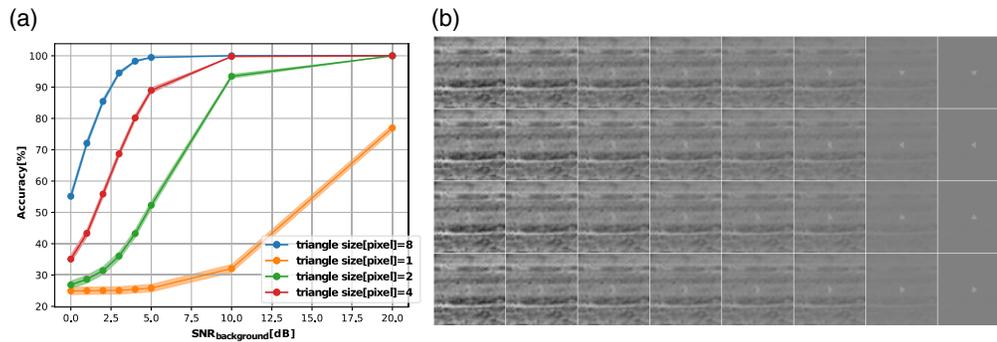


Fig. 5 (a) Accuracy over $\text{SNR}_{\text{background}}$ (dB): fixed $\text{SNR}_{\text{noise}}$ (dB) = 40 and varying triangle circumradius r . (b) Example validation images: triangle circumradius r (pixel) = 2, four orientations, and eight levels of $\text{SNR}_{\text{background}}$ (dB) $\in [0, 1, 2, 3, 4, 5, 10, 20]$.

3.2 Upscaling of Receptive Field

Models according to the CNN architecture shown in Fig. 2 were trained for various resolutions of receptive fields, i.e., 128×128 , 256×256 , 512×512 , and 1024×1024 . For 512×512 and 1024×1024 , a reduction of learning rate $\eta = 10^{-4}$ was required to achieve significant model improvements compared with the initial model states. Otherwise, model performances stagnated on average at 25% guessing rates. In Fig. 6, validation accuracies over $\text{SNR}_{\text{background}}$ are shown for different sizes of receptive fields.

There is a general trend of decreasing accuracies for larger receptive fields. This may be due to the fact that larger receptive fields can contain more objects with high similarity to triangles. Furthermore, the growth factor $g = 1.2$ for the number of feature maps per block $n_{\text{feature-maps}}$ may be insufficiently small to provide enough model capacity for an increasing range of triangle sizes to handle. A surprising fact is the lower accuracies for higher $\text{SNR}_{\text{noise}} = 20$ dB compared with $\text{SNR}_{\text{noise}} = 5$ dB and receptive fields of 256×256 pixels and larger. This might indicate beneficial properties of Gaussian noise by suppressing structures in the background resembling the triangle target.

3.3 Comparison of Methods for Contrast Enhancement

A trained 64×64 model was validated by images processed with the 25 ADSP algorithms shown in Table 1. Many algorithms only operate on integer pixel values based on gray level distributions, which are widespread for natural images. Hence, to prevent saturation due to clipping, the pixel values are shifted to have a mean of half of the dynamic range, upsampled by a factor $2^8 - 1 = 255$ and converted to 8 bit. After ADSP processing, this shifting and upscaling is reverted, and pixel values are converted to floating point numbers.

In Fig. 7, differences of accuracies between single CE algorithms and the identity are shown for varying $\text{SNR}_{\text{background}}$ and $\text{SNR}_{\text{noise}}$. For convenience, the algorithms were ranked with respect to the maximum value, and only the top 10 algorithms are shown.

It can be observed that there are accuracy gains for low $\text{SNR}_{\text{background}} < 5$ dB, and the accuracy differences are quite similar for the top 10 algorithms. Because the accuracy is about 100% for $\text{SNR}_{\text{background}} \geq 5$ dB and $\text{SNR}_{\text{noise}} \geq 5$ dB without ADSP processing according to Fig. 4, no significant improvement by CE can be expected for these cases. By contrast, a severe degradation of model performance occurs for some CE algorithms and high $\text{SNR}_{\text{noise}} = 20$ dB.

To validate CE algorithms on a variety of degradations, different triangle parameters and degradation parameters were varied and uniformly distributed in the ranges given in Table 2.

$N_{\text{target}} = 1000$ random samples of triangles and degradation parameters were combined with $N_{\text{backgrounds}} = 1000$ natural backgrounds as random crops from Open Images V6.⁸ This procedure was repeated $N_{\text{chunks}} = 10$ times with varying random seeds, resulting in different triangles and background images. Model accuracies were calculated on $N_{\text{total}} = N_{\text{chunks}} N_{\text{targets}} N_{\text{backgrounds}} = 10^7$ images with 64×64 pixels. The same procedure was

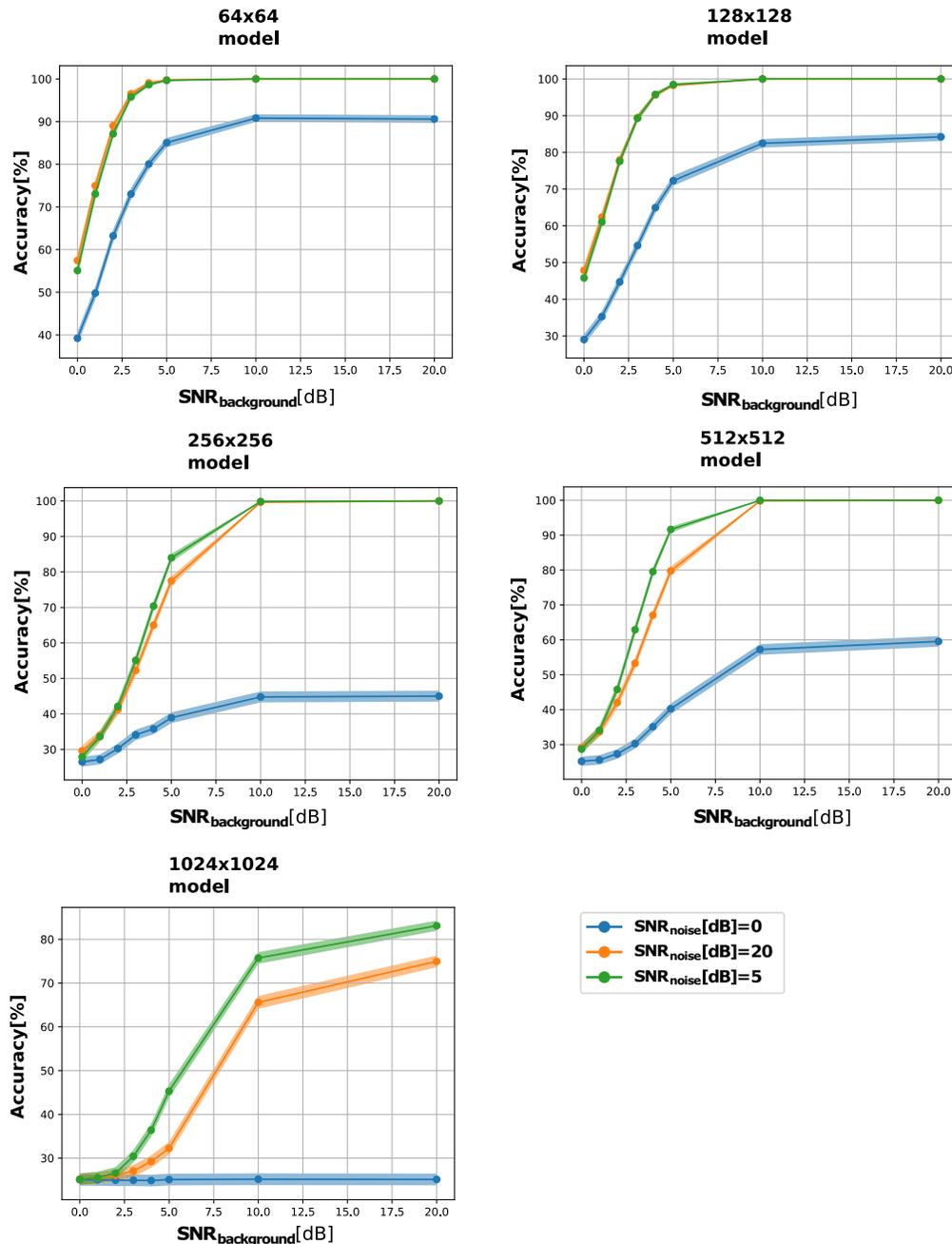


Fig. 6 Validation accuracy over $\text{SNR}_{\text{background}}$ (dB) for different sizes of receptive fields and $\text{SNR}_{\text{noise}} \in [0, 5, 20]$ dB based on $N_{\text{total}} = N_{\text{direction}} \cdot N_{\text{background}} = 4000$ images containing a centered triangle with circumradius $r = 10$ pixel, superposed by $N_{\text{background}} = 1000$ different backgrounds formed by random image crops of respective sizes from the OpenImages V6 test set and $N_{\text{direction}} = 4$ triangle orientations. For increasing field size, the model performance decreases.

repeated, with each of the 25 CE algorithms from Table 1 being applied respectively as the final step. Compared with grid variation of individual triangle and degradation parameters, random sampling of many of these parameters allows for investigation of individual parameters by arbitrary parameter cuts, whereas other parameters are widely distributed. This gives better insights on possible fluctuations on model performances when those parameters are unknown.

As shown in Fig. 8, accuracy differences between each of the 25 CE algorithms and identity were calculated for parameter cuts of the triangle circumradius r and the signal-to-background ratio $\text{SNR}_{\text{background}}$. Only accuracies on images with values in $r \in [1, 3]$ pixel (left),

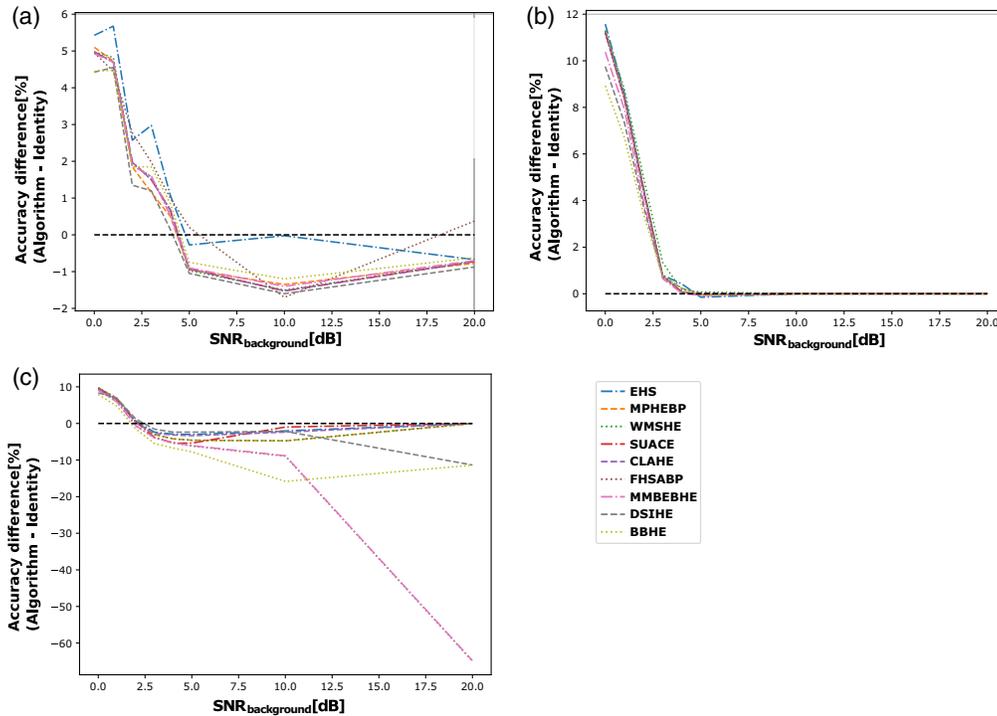


Fig. 7 Accuracy difference $\Delta I(\%) = I_{CE}(\%) - I_0(\%)$ for top 10 CE algorithms ranked with respect to the maximal value. Accuracies I_{CE} (I_0) for every data point are based on evaluations of the 64×64 TOD model on $N_{total} = N_{direction} \cdot N_{background} = 4000$ artificial images with (without) CE containing a centered triangle with circumradius $r = 10$ pixel, $N_{direction} = 4$ triangle orientations, and $N_{background} = 1000$ different backgrounds formed by random image crops from the OpenImages V6⁸ test set. In addition, before CE, Gaussian noise was added: (a) $SNR_{noise} = 0$ dB, (b) $SNR_{noise} = 5$ dB, and (c) $SNR_{noise} = 20$ dB.

Table 2 Boundaries for uniformly distributed triangle parameters and degradation parameters with the image dimensions $l = 64$ and the dynamic range $DR = 255$.

Parameter	Min	Max
Background level $c_{background}$	$0.25DR$	$0.75DR$
Amplitude $a_{triangle}$	$-0.25DR$	$0.25DR$
Triangle circumradius r	0	$0.25l$
Horizontal position	$0.25l$	$0.75l$
Vertical position	$0.25l$	$0.75l$
$SNR_{background}$ (dB)	0	20
SNR_{noise} (dB)	0	20
Orientation angle α (deg)	0	360

$r \in [14,16]$ pixel (right), $SNR_{background} \in [0,2]$ dB (left), and $SNR \in [18,20]$ dB (right) were selected. The interquartile ranges (IRQ), shown as orange boxes, contain values between the 25%-percentile and the 75%-percentile. The IRQs are extended by whiskers by $1.5IRQ$ at both sides at maximum, but they are limited by the respective minimal and maximal values in the data. Outliers are shown as circle markers.

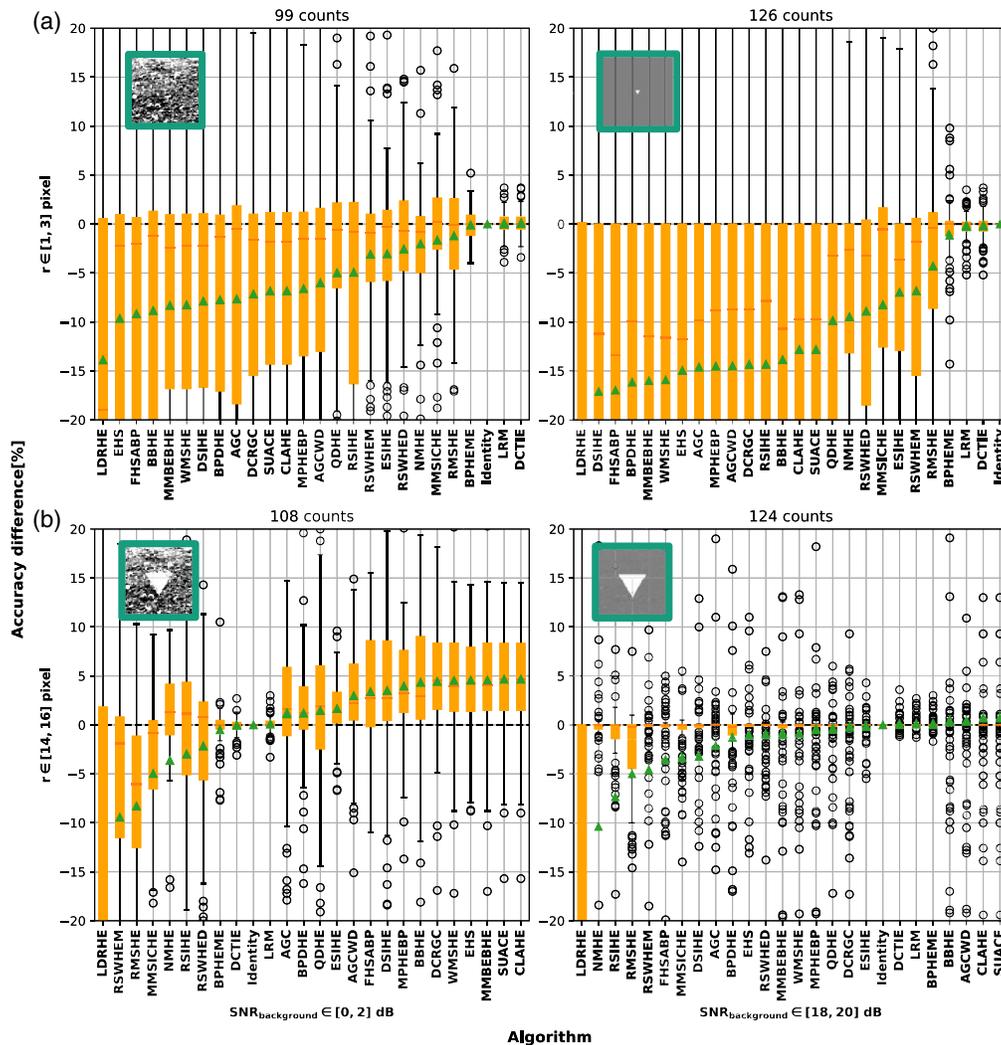


Fig. 8 Whisker-box plots of accuracy differences of the 64×64 model between 25 CE algorithms and identity based on varying cuts for (a) triangle circumradius $r \in [1,3]$ pixel, (b) $r \in [14,16]$ pixel, $\text{SNR}_{\text{background}} \in [0,2]$ dB (left), and $\text{SNR}_{\text{background}} \in [18,20]$ dB (right). Insets show examples of centered triangles for r , $\text{SNR}_{\text{background}}$ in the corresponding ranges with $\text{SNR}_{\text{noise}} = \infty$. For each algorithm, means (green triangle marker), medians (red lines), IQRs (orange boxes) of accuracy differences for the counts in the respective parameter intervals are shown (further details in the text). At high $\text{SNR}_{\text{background}}$ and low triangle circumradius r , the 25 CE algorithms lead to significant impairment of the model accuracies.

Obviously, a high $\text{SNR}_{\text{background}}$ and a low triangle circumradius r lead to significant impairment of model accuracies by most of the 25 CE algorithms. Accuracy differences at high triangle circumradius r and high $\text{SNR}_{\text{background}}$ (right bottom) show small IRQs, as the accuracy is often saturated at 100% for high $\text{SNR}_{\text{noise}}$. Hence, accuracies for low $\text{SNR}_{\text{noise}}$ are rendered as outliers. For high $r \in [14,16]$ pixel and low $\text{SNR}_{\text{background}} = [0,2]$ dB (left bottom) only, some CE algorithms show accuracy gains. Also, monotonic transitions of accuracy differences for varying r and $\text{SNR}_{\text{background}}$ were observed. The reason for the significant impairment at high $\text{SNR}_{\text{background}}$ could be due to the fact that a narrow gray level distribution of background values leads to excessive enhancement of the background by most CE algorithms, resulting in textures with a low dynamic range, steep edges, and a high similarity to the triangle to be discriminated. On the other hand, a large triangle reduces the number of background pixels and hence their contribution to the gray level distribution of the entire image. Most of the investigated CE algorithms depend on the image gray level distribution.

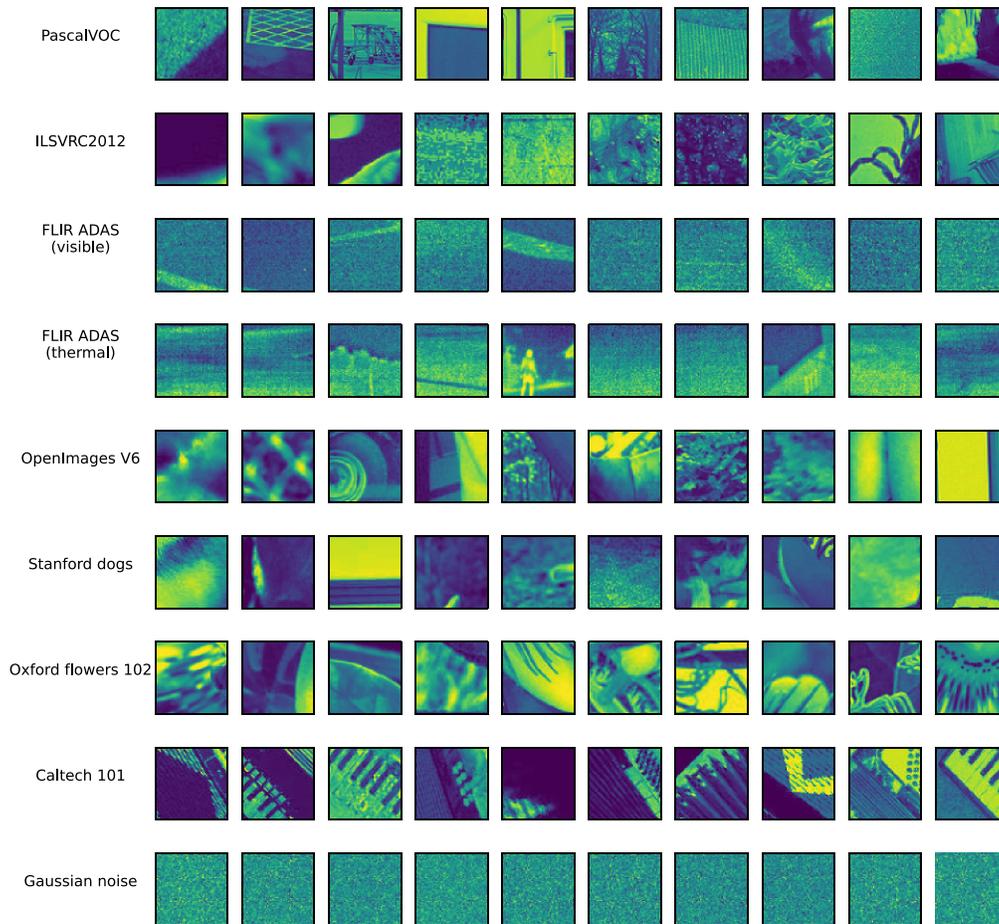


Fig. 9 Examples of 64×64 random crops of different image databases used as backgrounds: Pascal VOC,⁴⁸ ILSVRC2012,⁴⁹ FLIR ADAS,⁵⁰ OpenImages V6,⁸ Stanford dogs,⁵¹ Oxford flowers 102,⁵² Caltech 101,⁵³ and Gaussian noise. Image are converted to single-channel data by averaging over channels of RGB data. For clarity, example images are shown using the colormap “viridis,” and pixel values are scaled to have a mean that equals the center of the dynamic range and a standard deviation of one-fourth of the dynamic range. Compared with other image databases, FLIR ADAS contains images with a high noise content.

3.4 Generalization on Background Images of Different Image Databases

To investigate the ability of the TOD model to generalize to a larger variety of background images, the trained 64×64 model was validated on artificial images with a single centered triangle with a fixed circumradius of $r = 10$ pixel superposed with background images resulting from random crops of images from different image databases. Examples of such image crops with 64×64 pixels are shown in Fig. 9 for different image databases: Pascal VOC,⁴⁸ ILSVRC2012,⁴⁹ FLIR ADAS,⁵⁰ OpenImages V6,⁸ Stanford dogs,⁵¹ Oxford flowers 102,⁵² Caltech 101,⁵³ and Gaussian noise.

In Fig. 10, the model accuracies over $N = N_{\text{direction}} \cdot N_{\text{background}}$ images and different $\text{SNR}_{\text{background}}$ are shown, with $N_{\text{direction}} = 4$ triangle orientations and $N_{\text{background}} = 1000$ different backgrounds from several image databases. In addition, the generated artificial images are impaired by Gaussian noise with a high noise level $\text{SNR}_{\text{noise}} = 0$ dB (left) and a low noise level $\text{SNR}_{\text{noise}} = 20$ dB. It can be observed that model accuracies are very similar for most of the image databases. In contrast, background images of Gaussian noise yield significantly better accuracies than those of the image databases. Images of FLIR ADAS⁵⁰ show accuracies between those of Gaussian noise and the image databases, which may be due to the relatively high noise content in the FLIR ADAS images. This fact indicates that structured backgrounds from the

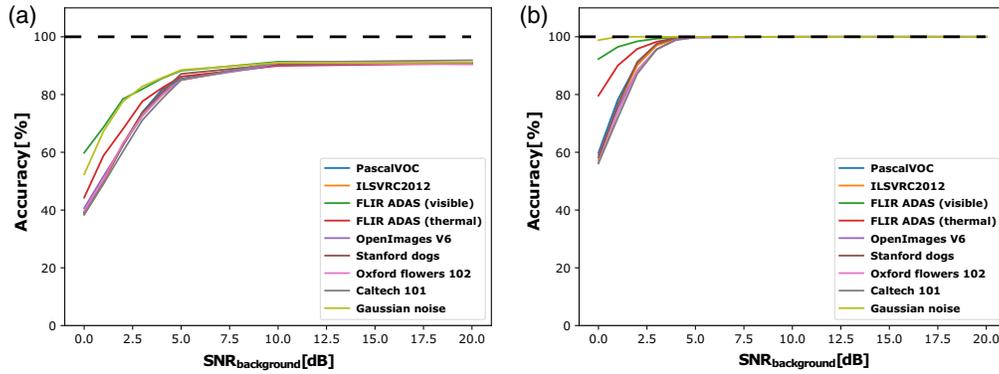


Fig. 10 Model accuracies over $\text{SNR}_{\text{background}}$ (dB). For each data point, $N = N_{\text{direction}} \cdot N_{\text{background}}$ artificial images are generated with $N_{\text{direction}} = 4$ triangle orientations and $N_{\text{background}} = 1000$ different backgrounds used from random crops of different image databases. Gaussian noise is added with (a) $\text{SNR}_{\text{noise}} = 0$ dB and (b) $\text{SNR}_{\text{noise}} = 20$ dB.

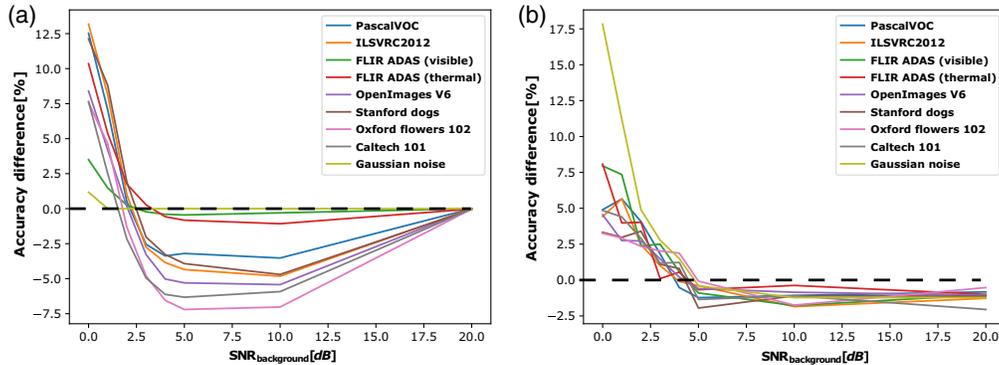


Fig. 11 Model accuracy differences $\Delta I = I_{\text{CLAHE}} - I_{\text{identity}}$ over $\text{SNR}_{\text{background}}$ (dB) with accuracies I_{CLAHE} based on images enhanced by CLAHE and accuracies I_{identity} for images without CE. For each data point, $N = N_{\text{direction}} \cdot N_{\text{background}}$ artificial images are generated with $N_{\text{direction}} = 4$ triangle orientations and $N_{\text{background}} = 1000$ different backgrounds used from random crops of different image databases. Gaussian noise is added with (a) $\text{SNR}_{\text{noise}} = 0$ dB and (b) $\text{SNR}_{\text{noise}} = 20$ dB.

image databases have a higher degradational effect on the triangle recognition than Gaussian noise for equal standard deviations of pixel value fluctuations.

The qualitative behavior of the model accuracy is similar for different databases when applying the methods for CE. The same is true for the ranges of $\text{SNR}_{\text{background}}$ values, for which CE yields an improvement in accuracy. For convenience, only an example for applying CLAHE is shown in Fig. 11.

3.5 Variation of Model Size

The used architectures so far were an arbitrary initial choice. One might ask if similar/better accuracies could have been achieved by smaller/larger models. To answer this question, further models were trained based on the default configuration (Sec. 2.6) with modifications of single parameters, i.e., the initial number of filters N_0 , the growth factor g , the dense layer size L , and the number of dense layers. In Fig. 12, validation accuracies for varying initial number of filters N_0 , the growth factor g , the dense layer size L , and the number of extra dense layers in addition to the final dense layer with four units are shown. Models are validated on images with $N_{\text{orientation}} = 4$ orientations, $N_{\text{background}} = 1000$, and $N_{\text{sample}} = 100$ samples, resulting in $N_{\text{total}} = N_{\text{orientations}} \cdot N_{\text{background}} \cdot N_{\text{sample}} = 400,000$ images. Accuracies denoted with “all degradations” contain images enhanced by one of the CE algorithms (Table 1) with equal probabilities.

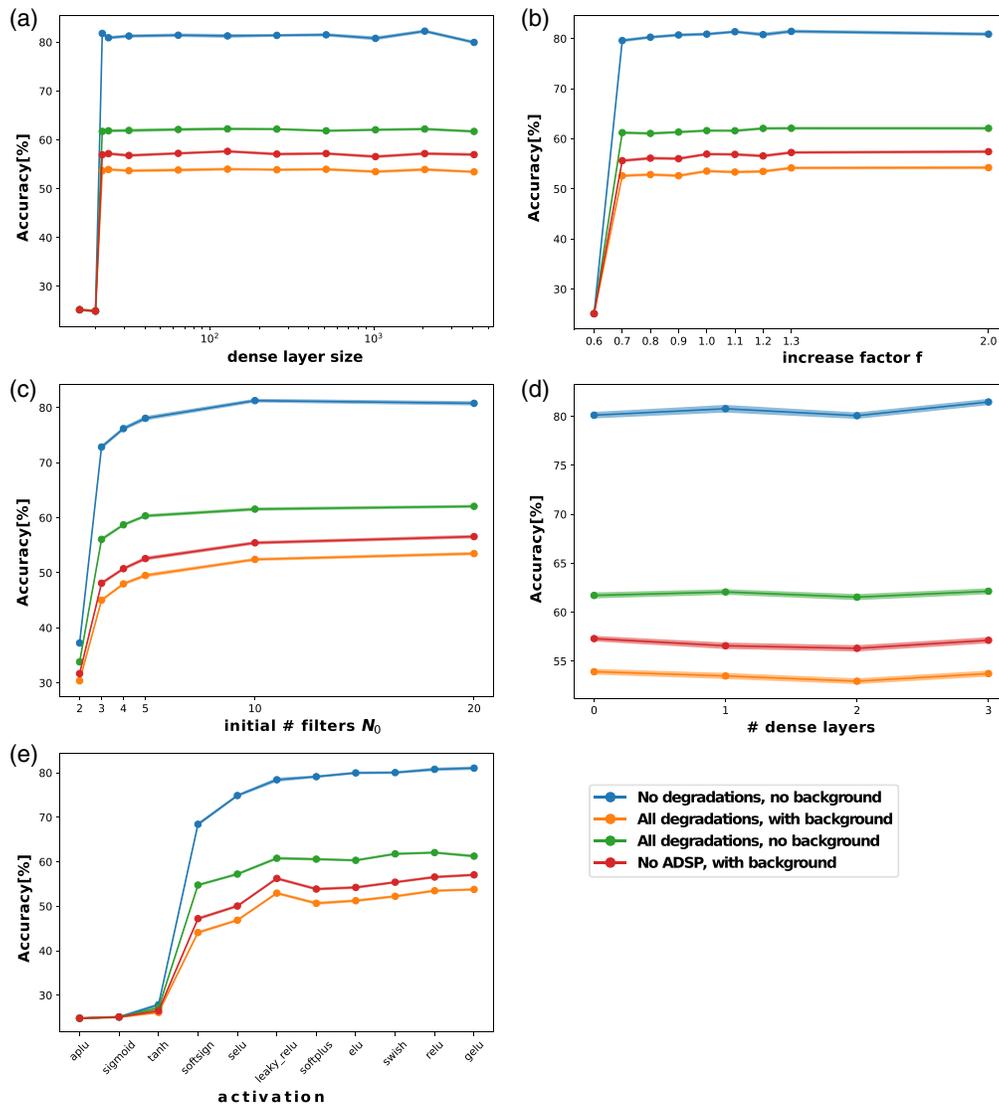


Fig. 12 Model comparisons for (a) varying dense layer size, (b) growth factor g , (c) initial number of filters N_0 , (d) the number of extra dense layers with $L = 1024$ units, and (e) different activation functions. Accuracies on different validation sets, each with $N = N_{\text{orientations}} \cdot N_{\text{backgrounds}} \cdot N_{\text{samples}} = 400,000$ images, with $N_{\text{backgrounds}} = 1000$, $N_{\text{orientations}} = 4$, $N_{\text{samples}} = 100$. “All degradations” include CE with one of the algorithms in Table 1.

Furthermore, validation accuracies are compared for different activations, i.e., leaky ReLU,⁵⁴ exponential linear unit (ELU),⁵⁵ Gaussian error linear unit (GELU),⁵⁶ scaled exponential linear unit (SELU),⁵⁷ APLU,⁵⁸ tanh, sigmoid,⁵⁹ softplus,⁵⁹ softsign,⁵⁹ and swish.⁶⁰ It can be observed that reductions of the dense layer size as low as $L = 22$, the growth factor $g = 0.7$, and the number of filters $N_0 = 10$ result in comparable accuracies compared with the default configuration. Even for a varying number of extra dense layers with $L = 1024$ units in addition to the final dense layer with four units, there are only slight variations in accuracies. From the 11 investigated activations, the ReLU⁵⁹ from the default configuration (Sec. 2.6) performs very well compared with most other activations. adaptive piecewise linear unit (APLU) and sigmoid did not converge at all above the guessing rate of 25%.

3.6 Model Complexity

We did benchmarks of our trained TOD model for 64x64 pixels on our machine (a Ryzen 9 3900X processor with an NVIDIA GeForce RTX 2080Ti graphics card and 64GB RAM).

Table 3 Model properties and benchmark results on an NVIDIA GeForce RTX 2080Ti.

Model	Input shape (height, width, # channels)	Parameters	FLOPs = 2*MACs (single image)	Average running time (ms) (batch size of 64 images, 100 calls, 95% confidence)
TOD	64 × 64 × 1	0.2M	0.2M	8.0 ± 0.2
TOD	128 × 128 × 1	0.3M	0.3M	8.9 ± 0.3
TOD	256 × 256 × 1	0.4M	0.4M	20 ± 4
TOD	512 × 512 × 1	0.5M	0.5M	80 ± 12
TOD	1024 × 1024 × 1	0.7M	0.7M	87 ± 13 (16 images)
VGG16 ⁶¹	224 × 224 × 3	138G	30.9G	110 ± 10
VGG19 ⁶¹	224 × 224 × 3	143G	39.2G	133 ± 11
ResNet50 v2 ⁶²	224 × 224 × 3	25.6G	7.96G	63.7 ± 1.5
Inception v3 ⁶³	299 × 299 × 3	23.8G	11.6G	105.8 ± 1.4
Xception ⁶⁴	299 × 299 × 3	22.9G	1.31G	162 ± 7

Table 3 gives the average running time on an NVIDIA GeForce RTX2080Ti, as well as the number of parameters and floating point operations (FLOPs), which equals twice the number of multiply-accumulate computations (MACs).

The trained TOD models are smaller and faster compared with current machine vision backbones, which are also shown in Table 3. Faster model inference allows for a stronger focus on several image degradations. Compared with the classification of RGB images in the visible spectrum, the TOD models are applicable on single-channel data, and the four triangle classes are symmetric and balanced. Furthermore, the triangle shape and texture are independent of any spectral band, in contrast to many image databases in the visible band. This is crucial, e.g., for range performance assessment of imagers in several infrared spectral bands [long-wavelength infrared (LWIR), mid-wavelength infrared (MWIR), and short-wavelength infrared (SWIR)].

3.7 Comparison with Other Image Quality Metrics

Different image quality metrics were proposed for assessment of methods for CE, such as absolute mean brightness error (AMBE),²⁰ discrete entropy,¹² measure of enhancement (EME) and EME based on entropy (EMEE),⁶⁵ QRCM,⁶⁶ UIQ,⁶⁷ EBCM,⁶⁸ and CII.²⁹ A more detailed overview of further image quality metrics and methods for CE can be found in another work.⁶⁹

However, most of these metrics were validated by subjective image quality assessments and may not correlate well with accuracies of models for TOD recognition or other machine vision tasks. To investigate some current nonreference metrics on images used in the evaluation of TOD models, these metrics were calculated for 64 × 64 images with a centered triangle superposed by backgrounds taken from OpenImages V6 scaled to different SNR_{background} and impaired by Gaussian noise with different SNR_{noise}. In addition, these images were enhanced by three CE methods, CLAHE,¹³ EHS,¹⁷ and SUACE,²⁸ which were among the top 10 algorithms in Fig. 7. In Fig. 13, values for nonreference image quality metrics EBCM,⁶⁸ EME, EMEE,⁶⁵ and entropy¹² over SNR_{background} are shown.

It can be observed that metric values Q_0 are high for low SNR_{background} and low SNR_{noise}, representing high variances of background and Gaussian noise, respectively. The metric values Q_0 decrease monotonically with increasing SNR_{background} and SNR_{noise}. The only exception is EBCM for SNR_{noise} = 20 dB, which we assume is due to the regularization of denominators in the algorithm. Therefore, low metric values represent conditions under which TOD accuracies are high. However, very similar metric values Q_0 could also be observed when the triangle was omitted. This indicates that these metrics are mainly determined by background for triangles

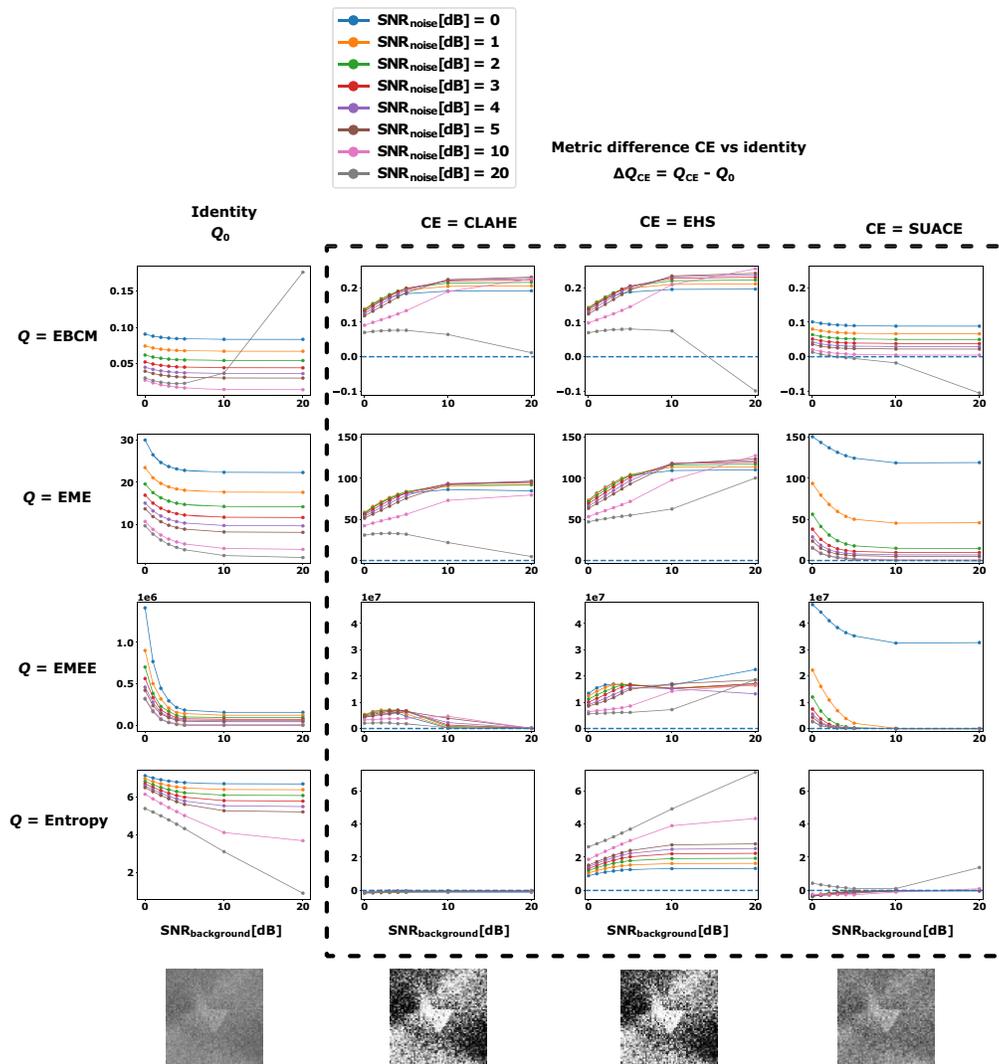


Fig. 13 Means of nonreference metric values Q_0 (EBCM,⁶⁸ EME,⁶⁵ EMEE,⁶⁵ and entropy,¹² left column) and metric differences $\Delta Q_{CE} = Q_{CE} - Q_0$ for three CE methods, i.e., CLAHE,¹³ EHS,¹⁷ and SUACE²⁸ (right columns) over $N_{\text{background}} = 1000$ images of 64×64 pixels with a centered triangle with circumradius $r = 10$ pixel for different SNR_{background}(dB) and SNR_{noise} (dB). $\Delta Q_{CE} = 0$ is shown as a horizontal blue dashed line. Example images with centered triangles superposed with background and impaired by Gaussian noise with SNR_{noise} = SNR_{background} = 3 dB (bottom), (left) without CE and (right) CE by the respective algorithm.

with $r \leq 10$ pixel. Thus, the metrics are weakly or not at all interrelated with the TOD task performance if the triangle is very small. CEs by the three CE methods generally lead to positive shifts of the metric values Q_{CE} or in other words $\Delta Q_{CE} = Q_{CE} - Q_0 > 0$, which can be interpreted as predominant enhancement of the background and noise, which aggravates TOD recognition. Similar results were found for the evaluation of the full-reference metrics AMBE,²⁰ CII,²⁹ QRCM,⁶⁶ and UIQ.⁶⁷ In summary, the metrics can be meaningful for the assessment of CE, whereas they cannot provide insights if the CE is beneficial for TOD recognition and possibly other classification tasks with small targets.

4 Conclusion

Accuracies of a sequential CNN model performing TOD discrimination were compared with respect to 25 different methods for CE. The background overlay was crucial because the accuracy is significantly impaired for high background variance and the CE algorithms strongly

depend on it. Accuracy gains for low signal-to-background ratios $\text{SNR}_{\text{background}} < 5$ dB and a sufficiently large triangle $r = 10$ pixel were shown. Model accuracies on images with randomly sampled triangle and degradation parameters revealed significant impairment by the investigated CE algorithms for a high $\text{SNR}_{\text{background}}$ and low triangle circumradius r . The strong fluctuations of accuracy differences highlight the difficulty in showing clear superiority of individual algorithms.

Models with increased resolution of the receptive field have shown decreasing accuracies, which may indicate that the growth of the number of model parameters was insufficient to represent the increasing range of triangle sizes. Another reason may be a higher number of background artifacts mimicking triangles. Larger images have more pixels. Therefore, their gray level distributions are statistically more stable. Hence, CE algorithms based on these gray level distributions should provide lower variations in the processed images and the associated accuracies. To prove this hypothesis, further investigations on larger receptive fields are required.

Variations of model size parameters, i.e., the number of filters N_0 , the growth factor g , the number of dense layers, and the activation function, have shown that the used default configuration is close to optimal based on the used model architecture and maximal values of degradation parameters used for the generation of training/validation data. Stronger degradations may require larger models for optimal accuracies.

The presented method can be used in an analogous way to assess the impact of other scene-based ADSP on military tasks. Moreover, the trained models can be used together with a test bed with an infrared scene projector for hardware in the loop testing of images including embedded ADSP. Finally, the methodology may be easily extended to more sophisticated classification tasks with real target signatures. In contrast to the triangle, real target signatures also have textures with spatial variations. Therefore, the gray level distribution and the CE based on it depend more strongly on the variations within the target, especially if the target covers lots of image pixels. Features related to these textures may require larger models compared with those investigated in this work.

References

1. P. Bijl and J. M. Valetton, "Triangle orientation discrimination: the alternative to minimum resolvable temperature difference and minimum resolvable contrast," *Opt. Eng.* **37**(7), 1976–1983 (1998).
2. S. Keßler et al., "The European computer model for optronic system performance prediction (ECOMOS)," *Proc. SPIE* **10433**, 282–294 (2017).
3. A. d'Acremont et al., "CNN-based target recognition and identification for infrared imaging in defense systems," *Sensors* **19**(9), 2040 (2019).
4. Y. Shi et al., "Scalable compression for machine and human vision tasks via multi-branch shared module," *J. Electron. Imaging* **31**, 023014 (2022).
5. Q. Wang, L. Shen, and Y. Shi, "Recognition-driven compressed image generation using semantic-prior information," *IEEE Signal Process. Lett.* **27**, 1150–1154 (2020).
6. D. Wegner and E. Repasi, "Imager assessment by classification of geometric primitives," *Proc. SPIE* **11406**, 17–25 (2020).
7. D. Wegner and E. Repasi, "Image based performance analysis of thermal imagers," *Proc. SPIE* **9820**, 982016 (2016).
8. A. Kuznetsova et al., "The Open Images Dataset V4: unified image classification, object detection, and visual relationship detection at scale," *Int. J. Comput. Vis.* **128**, 1956–1981 (2020).
9. S.-C. Huang, F.-C. Cheng, and Y.-S. Chiu, "Efficient contrast enhancement using adaptive gamma correction with weighting distribution," *IEEE Trans. Image Process.* **22**, 1032–1041 (2013).
10. Y.-T. Kim, "Contrast enhancement using brightness preserving bi-histogram equalization," *IEEE Trans. Consum. Electron.* **43**, 1–8 (1997).
11. H. Ibrahim and N. Kong, "Brightness preserving dynamic histogram equalization for image contrast enhancement," *IEEE Trans. Consum. Electron.* **53**, 1752–1758 (2007).

12. C. Wang and Z. Ye, "Brightness preserving histogram equalization with maximum entropy: a variational perspective," *IEEE Trans. Consum. Electron.* **51**, 1326–1334 (2005).
13. K. Zuiderveld, "Graphics Gems IV," in *Contrast Limited Adaptive Histogram Equalization*, pp. 474–485, Academic Press Professional, Inc., San Diego, California (1994).
14. Z.-G. Wang, Z.-H. Liang, and C.-L. Liu, "A real-time image processor with combining dynamic contrast ratio enhancement and inverse gamma correction for PDP," *Displays* **30**(3), 133–139 (2009).
15. J. Tang, E. Peli, and S. Acton, "Image enhancement using a contrast measure in the compressed domain," *IEEE Signal Process. Lett.* **10**, 289–292 (2003).
16. Y. Wang, Q. Chen, and B. Zhang, "Image enhancement based on equal area dualistic sub-image histogram equalization method," *IEEE Trans. Consum. Electron.* **45**, 68–75 (1999).
17. D. Coltuc, P. Bolon, and J.-M. Chassery, "Exact histogram specification," *IEEE Trans. Image Process.* **15**, 1143–1152 (2006).
18. S. F. Tan and N. A. M. Isa, "Exposure based multi-histogram equalization contrast enhancement for non-uniform illumination images," *IEEE Access* **7**, 70842–70861 (2019).
19. C. Wang, J. Peng, and Z. Ye, "Flattest histogram specification with accurate brightness preservation," *IET Image Process.* **2**, 249–262 (2008).
20. S.-D. Chen and A. Ramli, "Minimum mean brightness error bi-histogram equalization in contrast enhancement," *IEEE Trans. Consum. Electron.* **49**, 1310–1319 (2003).
21. K. Singh and R. Kapoor, "Image enhancement via median-mean based sub-image-clipped histogram equalization," *Optik* **125**(17), 4646–4651 (2014).
22. K. Wongsritong et al., "Contrast enhancement using multipeak histogram equalization with brightness preserving," in *IEEE Asia-Pacific Conf. Circuits and Syst., 1998. IEEE APCCAS 1998*, pp. 455–458 (1998).
23. S. Poddar et al., "Non-parametric modified histogram equalisation for contrast enhancement," *IET Image Process.* **7**, 641–652 (2013).
24. C. H. Ooi and N. A. M. Isa, "Adaptive contrast enhancement methods with brightness preserving," *IEEE Trans. Consum. Electron.* **56**, 2543–2551 (2010).
25. S.-D. Chen and A. Ramli, "Contrast enhancement using recursive mean-separate histogram equalization for scalable brightness preservation," *IEEE Trans. Consum. Electron.* **49**, 1301–1309 (2003).
26. K. Sim, C. Tso, and Y. Tan, "Recursive sub-image histogram equalization applied to gray scale images," *Pattern Recognit. Lett.* **28**(10), 1209–1221 (2007).
27. M. Kim and M. Chung, "Recursively separated and weighted histogram equalization for brightness preservation and contrast enhancement," *IEEE Trans. Consum. Electron.* **54**, 1389–1397 (2008).
28. A. M. R. R. Bandara, K. A. S. H. Kulathilake, and P. W. G. R. M. P. B. Giragama, "Super-efficient spatially adaptive contrast enhancement algorithm for superficial vein imaging," in *IEEE Int. Conf. Ind. and Inf. Syst. (ICIIS)*, pp. 1–6 (2017).
29. F. Bulut, "Low dynamic range histogram equalization (LDR-HE) via quantized Haar wavelet transform," *Visual Comput.* **38**(6), 2239–2255 (2022).
30. J. D. Fahnestock and R. A. Schowengerdt, "Spatially variant contrast enhancement using local range modification," *Opt. Eng.* **22**(3), 223378 (1983).
31. P.-C. Wu, F.-C. Cheng, and Y.-K. Chen, "A weighting mean-separated sub-histogram equalization for contrast enhancement," in *Int. Conf. Biomed. Eng. and Comput. Sci. (ICBECS)*, pp. 1–4 (2010).
32. D. Wegner and S. Keßler, "Comparison of algorithms for contrast enhancement based on TOD assessments by convolutional neural networks," *Proc. SPIE* **12271**, 122710H (2022).
33. J. Park et al., "Distort-and-recover: color enhancement using deep reinforcement learning," in *Proc. IEEE Conf. Comput. Vis. and Pattern Recognit. (CVPR)* (2018).
34. B. Xiao et al., "Histogram learning in image contrast enhancement," in *IEEE/CVF Conf. Comput. Vis. and Pattern Recognit. Workshops (CVPRW)*, pp. 1880–1889 (2019).
35. G. F. C. Campos et al., "Machine learning hyperparameter selection for contrast limited adaptive histogram equalization," *EURASIP J. Image Video Process.* **2019**(1), 59 (2019).
36. Y.-G. Shin et al., "Unsupervised deep contrast enhancement with power constraint for OLED displays," *IEEE Trans. Image Process.* **29**, 2834–2844 (2020).

37. V. Bychkovsky et al., “Learning photographic global tonal adjustment with a database of input/output image pairs,” in *CVPR*, IEEE (2022).
38. S. Lombardi and K. Nishino, “Reflectance and illumination recovery in the wild,” *IEEE Trans. Pattern Anal. Mach. Intell.* **38**(1), 129–141 (2016).
39. H. Yue et al., “Contrast enhancement based on intrinsic image decomposition,” *IEEE Trans. Image Process.* **26**(8), 3981–3994 (2017).
40. G. B. Airy, “On the diffraction of an object-glass with circular aperture,” *Trans. Cambridge Philos. Soc.* **5**, 283 (1835).
41. G. C. Holst, *Electro-Optical Imaging System Performance*, 6th ed., JCD Publishing (2017).
42. D. P. Kingma and J. Ba, “Adam: a method for stochastic optimization,” CoRR abs/1412.6980 (2014).
43. K. He et al., “Delving deep into rectifiers: surpassing human-level performance on ImageNet classification,” CoRR abs/1502.01852 (2015).
44. S. Ioffe and C. Szegedy, “Batch normalization: accelerating deep network training by reducing internal covariate shift,” in *Proc. 32nd Int. Conf. Mach. Learn.*, F. Bach and D. Blei, eds., PMLR, Lille, France, Vol. 37, pp. 448–456 (2015).
45. T. Salimans and D. P. Kingma, “Weight normalization: a simple reparameterization to accelerate training of deep neural networks,” CoRR abs/1602.07868 (2016).
46. A. Brock et al., “High-performance large-scale image recognition without normalization,” (2021).
47. A. Yazdanbakhsh et al., “An evaluation of edge TPU accelerators for convolutional neural networks,” CoRR abs/2102.10423 (2021).
48. M. Everingham et al., “The pascal visual object classes (VOC) challenge,” *Int. J. Comput. Vis.* **88**(2), 303–338 (2010).
49. O. Russakovsky et al., “ImageNet large scale visual recognition challenge,” *Int. J. Comput. Vis.* **115**(3), 211–252 (2015).
50. FLIR, “FLIR ADAS Teledyne,” <https://www.flir.de/oem/adas/adas-dataset-form/> (accessed on 2 August 2023).
51. A. Khosla et al., “Novel dataset for fine-grained image categorization,” in *First Workshop on Fine-Grained Visual Categorization, IEEE Conf. Comput. Vis. and Pattern Recognit.*, Colorado Springs, CO (2011).
52. M.-E. Nilsback and A. Zisserman, “Automated flower classification over a large number of classes,” in *Indian Conf. Comput. Vis., Graph. and Image Process.* (2008).
53. F.-F. Li et al., “Caltech 101,” (2022).
54. A. Maas, A. Hannun, and A. Ng, “Rectifier nonlinearities improve neural network acoustic models” (2013).
55. D. Clevert, T. Unterthiner, and S. Hochreiter, “Fast and accurate deep network learning by exponential linear units (ELUs),” CoRR abs/1511.07289 (2015).
56. D. Hendrycks and K. Gimpel, “Bridging nonlinearities and stochastic regularizers with Gaussian error linear units,” CoRR abs/1606.08415 (2016).
57. G. Klambauer et al., “Self-normalizing neural networks,” CoRR abs/1706.02515 (2017).
58. F. Agostinelli et al., “Learning activation functions to improve deep neural networks,” CoRR abs/1412.6830 (2014).
59. Tensorflow, “Tensorflow API,” (2022). https://www.tensorflow.org/api_docs/python/tf/keras/activations.
60. P. Ramachandran, B. Zoph, and Q. V. Le, “Searching for activation functions,” CoRR abs/1710.05941 (2017).
61. K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” CoRR abs/1409.1556 (2014).
62. K. He et al., “Identity mappings in deep residual networks,” Technical Report (2016). arXiv:1603.05027 [cs].
63. C. Szegedy et al., “Rethinking the inception architecture for computer vision,” CoRR abs/1512.00567 (2015).
64. F. Chollet, “Xception: deep learning with depthwise separable convolutions,” CoRR abs/1610.02357 (2016).

65. S. Aghaian, B. Silver, and K. Panetta, "Transform coefficient histogram-based image enhancement algorithms using contrast entropy," *IEEE Trans. Image Process.* **16**, 741–758 (2007).
66. T. Celik, "Spatial mutual information and pagerank-based contrast enhancement and quality-aware relative contrast measure," *IEEE Trans. Image Process.* **25**(10), 4719–4728 (2016).
67. Z. Wang and A. Bovik, "A universal image quality index," *IEEE Signal Process. Lett.* **9**, 81–84 (2002).
68. B. N. Anoop, P. E. Ameenudeen, and J. Joseph, "A meta-analysis of contrast measures used for the performance evaluation of histogram equalization based image enhancement techniques," in *9th Int. Conf. Comput., Commun. and Netw. Technol. (ICCCNT)*, pp. 1–6 (2018).
69. S. V. Renuka, D. R. Edla, and J. Joseph, "An objective measure for assessing the quality of contrast enhancement on magnetic resonance images," *J. King Saud Univ. – Comput. Inf. Sci.* **34**(10, Part B), 9732–9744 (2021).

Daniel Wegner is a research assistant at the Fraunhofer Institute of Optronics, System Technologies and Image Exploitation (IOSB) in Ettlingen, Germany. He received his diploma degree in physics (equivalent to MS degree) from the Karlsruhe Institute of Technology (KIT), Germany, in 2013. Then, he received his PhD in physics from KIT in 2022. His research interests include image quality metrics, methods for contrast enhancement, image-based simulation of atmospheric turbulence as well as approaches for modeling, image-based simulation, and machine learning for imager performance assessment.

Stefan Keßler is the head of the Sensor Simulation Group at the Fraunhofer Institute of Optronics, System Technologies and Image Exploitation in Ettlingen, Germany. He received his diploma degree in physics from the University of Heidelberg in 2008 and his PhD in physics from the University of Erlangen-Nürnberg in 2014. His research activities comprise sensor modeling, image simulation of infrared and electro-optical imagers, and imager performance assessment.