How Does Image Quality Affect Computer-Aided Diagnosis of Colorectal Polyps?

How Does Image Quality Affect Computer-Aided Diagnosis of Colorectal Polyps?

Thom Scheeve[a], Nikoo Dehghani[a], Quirine E.W. van der Zander[b], Ayla Thijssen[b], Ramon-Michel Schreuder[c], Ad A.M. Masclee[b], Erik J. Schoon[c], Fons van der Sommen[a], and Peter H.N. de With[a]

[a]Eindhoven University of Technology, P.O. Box 513, Eindhoven 5600 MB, the Netherlands
[b]Maastricht University Medical Center+, P.O. Box 5800, Maastricht 6202 AZ, the Netherlands
[c]Catharina Hospital, P.O. Box 1350, Eindhoven 5602 ZA, the Netherlands

ABSTRACT

Colorectal cancer (CRC) is one of the leading causes of cancer-related deaths with rising incidence. Since the survival rate of CRC is correlated with the cancer stage at diagnosis, timely detection and adequate treatment strategies are of utmost importance. Technical innovations such as machine learning (ML) and its application in endoscopy show promising results, but the trust of medical doctors in ML is lacking and the 'black box' nature complicates the understanding of such systems in clinical practice. In contrast to CT and MRI, image quality is a limiting factor in especially endoscopic imaging, as it is very operator dependent. However, the influence of image quality on convolutional (deep) neural networks (CNNs) is insufficiently studied in relation to clinical practice and the usage of medical image data for computer-aided detection and diagnosis (CADx) systems. This paper explores the influence of degraded image quality on the performance of CNNs applied to colorectal polyp (CRP) characterization. Five commonly used CNN architectures, from simple to more complex, are employed with a custom classification head for common CRP characterization. To degrade the quality of images, distortions such as noise, blur, and contrast changes are imposed on the data and their influence on the performance degradation is studied for the mentioned CNN architectures. A large prospectively collected *in vivo* data set, gathered from four Dutch, both academic and community, hospitals is employed. Results for CRP characterization show that promising CNN-based methods are rather susceptible to noise and blur distortions but reasonably resilient to changes in contrast. This implies that image quality needs monitoring and control prior to directly using image data in CNN models, in order to gain trustworthy use of deep learning (DL) models in a clinical setting. We propose that incorporating an image quality indicator in CADx systems will lead to better acceptance of such systems, and is necessary for the safe implementation of DL applications in clinical practice.

Keywords: Deep learning, image quality, image classification, reproducibility, colorectal cancer, gastroenterology, endoscopic imaging

1. INTRODUCTION

Cancer is the leading cause of death worldwide.[1] Colorectal cancer (CRC) remains the third leading cause in terms of incidence, but ranks second in terms of deaths worldwide, accounting for about one in ten cancer cases (10.0%; 1.9 million) and deaths (9.4%; 935,000) in 2020.[2] As with most cancer types, the survival rate of CRC is correlated with the cancer stage at diagnosis.[3,4] Precursor lesions of CRC are colorectal polyps (CRPs), which consist of two major classes: conventional adenomas (ADs) and serrated lesions. The most common serrated lesions are the hyperplastic polyps (HPs) and the sessile serrated lesions (SSLs). All ADs are dysplastic, or precancerous, whereas all HPs and a majority of SSLs are non-dysplastic.[5] Although SSLs have a low incidence, they can progress into cancer, while they are the hardest to detect because of their flat appearance. Therefore, timely detection and adequate treatment strategies are of utmost importance.

Further author information: (Send correspondence to Th.S.)
Th.S.: E-mail: t.scheeve@tue.nl

Medical Imaging 2023: Computer-Aided Diagnosis, edited by Khan M. Iftekharuddin, Weijie Chen, Proc. of SPIE Vol. 12465, 124650S · © 2023 SPIE
1605-7422 · doi: 10.1117/12.2654167

Technical innovations, such as computer-aided detection (CADe) and diagnosis (CADx) systems, are not yet (widely) adopted in clinical practice. This is caused by the lack of trust of medical doctors (MDs) in these systems, because their 'black box' nature complicates their understanding, which further hampers in relying on the outcomes of such systems. Despite the high intrinsic quality of modern endoscopic equipment, one limiting factor for endoscopic imaging is the continuous capturing with high image quality, which, in contrast to CT and MRI, is strongly depending on the capabilities of the endoscopist. This aspect may have severe implications on clinical care. For example, subtle flat lesions in the colon are difficult to detect endoscopically, since they are associated with less apparent morphological changes than other CRPs. Artifacts such as blurring due to motility, breathing and patient movement may lead to missed lesions and, as a result, can progress to cancer.

A large number of studies have been conducted on the application of machine learning (ML) in endoscopy, showing promising results for a variety of applications, ranging from CRP detection to esophageal cancer localization.[6–9] However, little research has been conducted on the influence of fluctuating image quality within endoscopic imaging due to capturing quality variations caused by the endoscopists on the performance of such promising automated ML algorithms. The quality variations can originate from poor system settings like suboptimal contrast of the video signal, or may be caused by the endoscopist during the procedure, when, *e.g.*, the endoscope is moved too quick while capturing leading to blur. Although some research on understanding how image quality affects ML can be found in literature,[10–12] there is no specific link made with clinical practice and medical image data and the impact on diagnostic applications remains unclear. Therefore, in this work, we study the influence of endoscopic image capturing quality on the performance of computer-aided diagnosis for CRP characterization.

The contributions of this paper are as follows. First, we present a framework to objectively evaluate the influence of several image quality parameters or distortions. Second, the CADx performance degradations are measured for noise, blur, and contrast. Third, we show that even network models that claim to be robust against such distortions are susceptible to intrinsic quality degradations, so that their usage in clinical practice should be accepted only when care is take on quality control.

2. METHODS

2.1 Study Sample

Endoscopic data is prospectively collected *in vivo* from patients who underwent a colonoscopy at the Catharina Hospital (CZE; Eindhoven, the Netherlands), Maastricht University Medical Center+ (MUMC+; Maastricht, the Netherlands), Bernhoven Hospital (BU; Uden, the Netherlands), and Zuyderland Medical Center (ZUY; Heerlen & Sittard-Geleen, the Netherlands) as of October 2017. For each patient, each CRP is represented by a number of image series captured from almost exactly the same position and distance to the polyp, but with different imaging conditions and enhancements. These conditions are depending on the applied brand of endoscope, giving the following types of images: one high-definition white light image (HDWL), two optically enhanced images (blue-light imaging and linked-color imaging; Fujifilm Corporation, Tokyo, Japan) without magnification, three optically enhanced images (i-Scan Modes 1, 2 and 3; Pentax Corporation, Tokio, Japan) without magnification, and one optically enhanced image in narrow-band imaging mode (Olympus Corporation, Tokio, Japan) without magnification. CRPs are divided in two categories using histopathology as gold standard: benign (HPs) and (pre)malignant (ADs, SSLs, and T1 colorectal carcinomas [T1-CRCs]).

2.2 Endoscopic Image Data Sets

As HDWL is the standard imaging modality in endoscopic imaging and available in all hospitals, only HDWL images are used in this study. In addition, by focusing on this basic modality only, we can study the influence of image quality in its purest form, without any possible influence of optical enhancements. Most studies on both CRP detection and characterization split CRPs in categories such as benign or premalignant, non-neoplastic or neoplastic, and hyperplastic polyp or adenoma, and therefore only include the HP and AD subtypes. As a consequence, HPs and ADs are included only in the data set from the study sample. The number of included HPs and ADs reflects their respective prevalence. Data is split into a training and test set such that the percentage of samples for each class is preserved in each split (stratified splits). Further details on the amount of images are given in Section 3 as they are coupled to the CRP types.

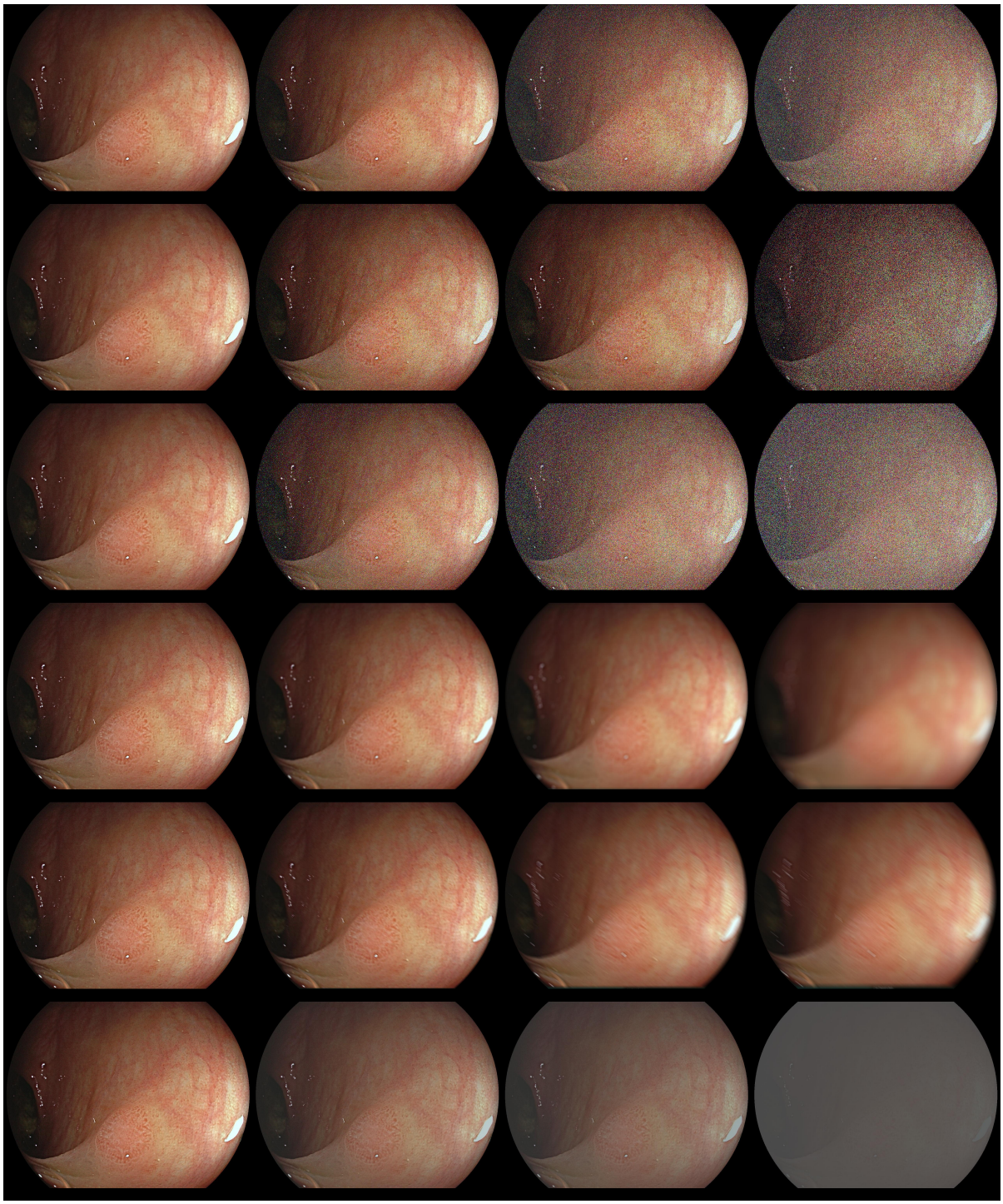

Figure 1: Endoscopic image of an adenoma, acquired with HDWL with addition of different quality distortions.
First Row: Original (left), Gaussian noise with $\sigma = 0.1$ (2^{nd} left), $\sigma = 0.3$ (2^{nd} right), and $\sigma = 0.5$ (right).
Second Row: Original (left), shot noise with $\lambda = 15$ (2^{nd} left), $\lambda = 10$ (2^{nd} right), and $\lambda = 1$ (right).
Third Row: Original (left), impulse noise with $p = 0.1$ (2^{nd} left), $p = 0.3$ (2^{nd} right), and $p = 0.5$ (right).
Fourth Row: Original (left), defocus blur with $\sigma = 5$ (2^{nd} left), $\sigma = 11$ (2^{nd} right), and $\sigma = 31$ (right).
Fifth Row: Original (left), motion blur with $k = 11$ (2^{nd} left), $k = 31$ (2^{nd} right), and $k = 51$ (right).
Sixth Row: Original (left), contrast reduced with $l = 0.7$ (2^{nd} left), $l = 0.5$ (2^{nd} right), and $l = 0.1$ (right).

Table 1: Distortion parameters and used values.

Distortion	Parameters	Values
Noise		
Gaussian	Kernel standard deviation ($\sigma \in \mathbb{R}_0$)	$\sigma = \{0, 0.1, \ldots, 1\}$
Shot	Rate ($\lambda \in \mathbb{R}_{>0}$)	$\lambda = \{10, 9, \ldots, 0\}$
Impulse	Probability ($p \in [0, 1]$)	$p = \{0, 0.1, \ldots, 1\}$
Blur		
Defocus	Kernel radius ($r \in \mathbb{R}_0$) Kernel standard deviation ($\sigma \in \mathbb{R}_0$)	$r = \{0, 1, \ldots, 10\}$ $\sigma = 1$
Motion	Kernel size ($k \in \mathbb{R}_{>0}$) Kernel standard deviation ($\sigma \in \mathbb{R}_0$) Rotation angle ($\alpha \in \mathbb{R}$)	$k = \{3, 9, 15, 21\}$ $\sigma = f(k) = \{k/3, 2k/3, k\}$ $\alpha = 0$
Contrast	Blending factor ($l \in \mathbb{R}_0$)	$l = \{2, 1.8, \ldots, 0\}$

2.3 Deep Learning Framework

Five commonly used convolutional (deep) neural network (CNN) architectures, from simple to more complex, are employed as backbone. We consider, more specifically, (Goog)LeNet, AlexNet, VGG-16, ResNet-50, and EfficientNet-B4, because these networks are typically employed in state-of-the-art CRP characterization algorithms. Each architecture is initialized with ImageNet(V1) pre-trained weights. Subsequently, for each architecture, the default classification head is replaced by a customized version. This replacement makes the resulting architectures suitable for our binary, histopathological classification problem. The custom classification head consists of two fully-connected layers with 1,024 neurons with a rectified linear unit (ReLU) activation function, followed by a fully-connected output layer with one neuron and a sigmoid activation function. The resulting neural networks are subsequently fine-tuned with the training data. Test time augmentation is employed to introduce various (deterministic) methods to degrade the quality of the test data and evaluate the influence on the performance of each neural network. The performance of each neural network is assessed with the area under the receiver operating characteristic curve (AUC) as figure of merit.

Reproducibility of studies employing CNNs is influenced by the order of the input during training procedures. In order to avoid these dependencies, we aim at training the various networks in the same way. For example, initialization of networks is usually performed in a random fashion and input data batches are being randomly shuffled in between training epochs. Since this randomness may hamper our ablation study on image quality, we opt to use the PyTorch Lightning framework (v1.8.6, Lightning AI), which features straightforward control of the training procedure.

2.4 Applied Distortions/Image Quality Degradation

Three types of common image quality distortions in endoscopic imaging are considered, namely: noise, blur and contrast changes (Figure 1). We consider three types of noise: noise due to low-lighting conditions, noise resulting from the light capturing process on the sensor, and noise caused by bit errors originated by faults in the imaging equipment. We briefly discuss these noise types below.

We model noise due to low-lighting conditions as additive Gaussian noise occurring in each color component of individual pixels. Noise from the capturing process is modeled as shot noise, also known as Poisson noise. As an analogue of salt-and-pepper noise, its color variation form is modeled as impulse noise, employed to model noise caused by bit errors.

Blur can occur due to movement of the endoscope, or the bowel, or when the endoscope is not focused properly on the CRP. We consider two types of blur: defocus blur and motion blur. Reduction in contrast can occur due to poor illumination of the CRP or an unclean endoscope lens. We obtain contrast reduction by blending the input images with a gray image. Parameters and corresponding values per distortion are shown in Table 1.

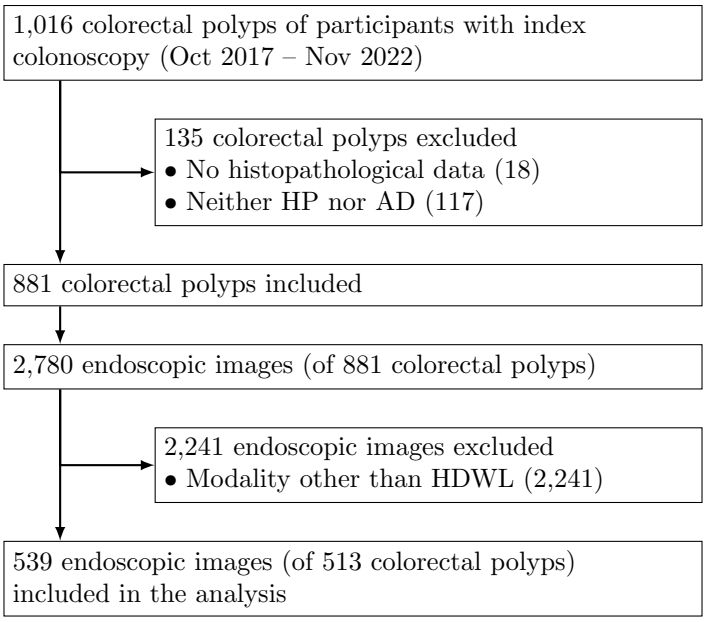

Figure 2: Flowchart of the included polyps and endoscopic images.

3. RESULTS

3.1 Colorectal Polyp Characteristics

Between October 2017 and November 2022, 1,016 CPRs were included in the COMET-OPTICAL research project. Histopathological data was required to be available ($n = 18$ missing) and CRPs needed to be either an AD or HP ($n = 117$ other histology), resulting in 881 CRPs that were included and a total of 2,780 endoscopic images. A total of 2,241 images were excluded, due to being captured in another modality than HDWL, resulting in 539 images of 513 CRPs eligible for analysis (Figure 2). Of the 513 CRPs, 95 were at least 10 mm, 74 had a size of 6–9 mm, and 342 were diminutive (less than or equal to 5 mm). Endoscopic characteristics of these CRPs are shown in Table 2. The median size of polyps was 4 mm, where 43% of the CRPs were located in the rectosigmoid ($n = 219$). Histology showed that 78% were adenomas and 22% hyperplastic polyps.

3.2 Influence of Distortions

Figure 3 shows the results of the influence of six distortions on the classification performance of five commonly used CNNs. One of the first observations is that all networks follow the shape of degradation which varies depending on the type of distortion and all networks decrease in performance with increasing level of degradation.

3.2.1 Noise

From Figures 3a–c, it is verily clear that the networks are susceptible to noise. Even for moderate noise levels (after one or two levels), the AUC decreases significantly. Being the most simple architecture, AlexNet is the most susceptible to any variant of noise. Even after one level of distortion growth, the AUC reaches values which indicate that the network is randomly guessing.

3.2.2 Blur

The networks are also susceptible to blur: the AUC decreases considerably for moderate blur levels (Figures 3d–e). Compared to motion blur, defocus blur seems to be more of a problem for the evaluated networks. As blur removes textures in images, important information will be lost which will affect the performance greatly in endoscopic imaging. Some CRPs are associated with less morphological changes than others. In addition, some CRPs may resemble the normal colonic mucosa, making them hard to detect, let alone correctly characterize.

Table 2: Endoscopic and histopathologic characteristics of adenomas and hyperplastic polyps included in the study.

	AD	HP
Number of lesions	402	111
Polyp size in mm, mean (SD)	7 (8)	4 (3)
Polyp size, n (%)		
1–2 mm	58 (14)	19 (17)
3–5 mm	196 (49)	69 (62)
6–9 mm	59 (15)	15 (14)
≥10 mm	87 (22)	8 (7)
Unclassified	2 (0)	– (–)
Location, n (%)		
Rectum	52 (13)	24 (22)
Sigmoid colon	108 (27)	35 (32)
Descending colon	52 (13)	4 (3)
Splenic flexure	2 (0)	– (–)
Transverse colon	67 (17)	19 (17)
Hepatic flexure	2 (0)	– (–)
Ascending colon	78 (20)	19 (17)
Caecum	41 (10)	10 (9)
Paris classification, n (%)		
Ip	16 (4)	2 (2)
Is	164 (41)	46 (41)
IIa	49 (12)	25 (23)
IIb	2 (0)	2 (2)
Unclassified	171 (43)	36 (32)
Dysplasia, n (%)		
Low-grade dysplasia	392 (98)	
High-grade dysplasia	10 (2)	

3.2.3 Contrast

The evaluated networks are surprisingly resilient to changes in contrast (Figure 3f). Since changes in contrast make textures less clear, we would expect, as for blur, loss of important information. In endoscopic imaging, texture is of great importance. As expected, the performance grows somewhat when the contrast increases to the reference level (level 0). When the contrast is deviating from the reference level, the performance rapidly decreases when lowering the contrast towards darkness.

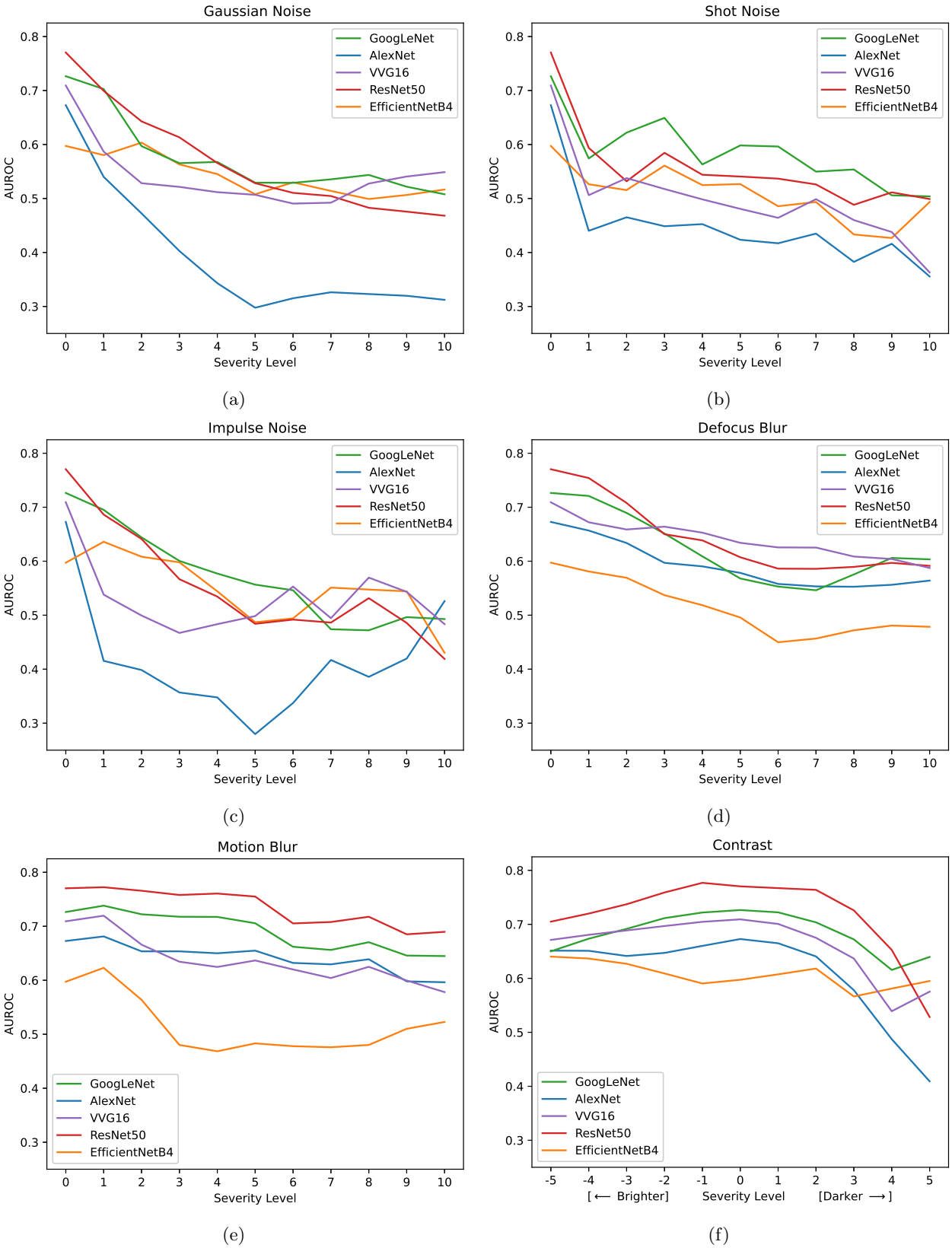

Figure 3: AUC performance degradations as a function of different types of quality distortions.

4. CONCLUSIONS

To the best of our knowledge, we are the first to investigate several common sources of image distortion in endoscopy and its impact on the performance of state-of-the-art computer-aided CRP characterization. To this end, we compared five commonly used CNN architectures for CRP characterization, as a function of the imposed image quality degradation or distortion.

The common finding for all networks is that their performance degrades with the growth of the distortion and the shape of the degradation curve seems to be similar for all networks. The simplest networks show the most degradation (*e.g.*, AlexNet), whereas a more complex architecture (*e.g.*, ResNet-50) is more robust against degradations. Results show that promising CNN-based methods for CRP characterization, are very susceptible to noise and blur, but surprisingly resilient to contrast changes. Also for most networks, the robustness for motion blur is quite good, which can be explained from the fact blur is concentrated around the region of interest (close to the polyps). Since most networks employ downscaling, the effect of motion blur is partly cancelled out due to the subsampling stages.

In endoscopic imaging in gastroenterology, image quality is of high importance for a robust and reliable application of deep learning models which are typically employed in CADe and CADx systems. The conducted experiments show that modern CNN architectures are not as resilient and robust against quality degradations as expected. This makes the use of these models critical when the circumstances are not controlled, especially in the medical field. We propose that incorporating an image quality indicator in CADx systems is a useful and necessary tool to assess whether the intrinsic quality of images is suited for automated, computer-aided analysis. When facilitating a good image quality control, CADe and CADx systems become more robust, which can lead to more trust in computer-aided systems and can help to more broadly accept and implement applications of deep learning models in clinical practice.

ACKNOWLEDGMENTS

We greatfully acknowledge the KWF Dutch Cancer Society for funding the COMET-OPTICAL project (Project No. 12639), which guided the presented study. The KWF Dutch Cancer Society did not contribute to the study protocol, data analysis or manuscript writing.

REFERENCES

[1] Bray, F., Laversanne, M., Weiderpass, E., and Soerjomataram, I., "The ever-increasing importance of cancer as a leading cause of premature death worldwide," *Cancer* **127**(16), 3029–3030 (2021).

[2] Sung, H., Ferlay, J., Siegel, R. L., et al., "Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries," *CA: A Cancer Journal for Clinicians* **71**(3), 209–249 (2021).

[3] Cohen, M. B., "Morphology of Cancer Precursor Lesions," in [*Cancer Precursors: Epidemiology, Detection, and Prevention*], Franco, E. L. and Rohan, T. E., eds., 20–29, Springer New York, New York, NY (2002).

[4] Rutter, C. M., Kim, J. J., Meester, R. G., et al., "Effect of Time to Diagnostic Testing for Breast, Cervical, and Colorectal Cancer Screening Abnormalities on Screening Efficacy: A Modeling Study," *Cancer Epidemiology, Biomarkers & Prevention* **27**, 158–164 (Feb. 2018).

[5] Rex, D. K., Hassan, C., and Bourke, M. J., "The colonoscopist's guide to the vocabulary of colorectal neoplasia: histology, morphology, and management," *Gastrointestinal Endoscopy* **86**, 253–263 (Aug. 2017).

[6] Mori, Y., Kudo, S.-e., Misawa, M., et al., "Real-Time Use of Artificial Intelligence in Identification of Diminutive Polyps During Colonoscopy," *Ann Intern Med* **169**, 357–366 (Sept. 2018). Publisher: American College of Physicians.

[7] Byrne, M. F., Chapados, N., Soudan, F., et al., "Real-time differentiation of adenomatous and hyperplastic diminutive colorectal polyps during analysis of unaltered videos of standard colonoscopy using a deep learning model," *Gut* **68**, 94–100 (Jan. 2019). Publisher: BMJ Publishing Group Section: Endoscopy.

[8] de Groof, J., van der Sommen, F., van der Putten, J., et al., "The Argos project: The development of a computer-aided detection system to improve detection of Barrett's neoplasia on white light endoscopy," *United European Gastroenterology Journal* **7**(4), 538–547 (2019).

[9] van der Zander, Q. E. W., Schreuder, R. M., Fonollà, R., et al., "Optical diagnosis of colorectal polyp images using a newly developed computer-aided diagnosis system (CADx) compared with intuitive optical diagnosis," *Endoscopy* **53**, 1219–1226 (Dec. 2021). Publisher: Georg Thieme Verlag KG.

[10] Dodge, S. and Karam, L., "Understanding How Image Quality Affects Deep Neural Networks," (Apr. 2016). Number: arXiv:1604.04004 arXiv:1604.04004 [cs].

[11] Su, J., Vargas, D. V., and Kouichi, S., "One pixel attack for fooling deep neural networks," *IEEE Trans. Evol. Computat.* **23**, 828–841 (Oct. 2019). arXiv:1710.08864 [cs, stat].
[12] Hendrycks, D. and Dietterich, T., "Benchmarking Neural Network Robustness to Common Corruptions and Perturbations," (Mar. 2019). arXiv:1903.12261 [cs, stat].