LONDON imaging MEETING 2021

IMAGING FOR DEEP LEARNING

SEPTEMBER 20 – 22 ONLINE

imaging.org

Sponsored by the Society for Imaging Science and Technology

FINAL PROGRAM AND PROCEEDINGS

The papers in this volume represent the program of the London Imaging Meeting 2021
held online September 20 – 22, 2021.

IS&T: Society for Imaging Science and Technology
7003 Kilworth Lane
Springfield, VA 22151 USA
703/642-9090; 703/642-9094 fax
info@imaging.org
www.imaging.org

ISBN: 978-0-89208-355-8
ISSN Online: 2694-118X
DOI: https://doi.org/10.2352/issn.2694-118X.2021.LIM-A
Contributions are reproduced from copy submitted by authors; no editorial changes have been made.

Funding for this meeting is supported by EPSRC.

WELCOME TO LIM 2021!

Thank you for joining us at this second London Imaging Meeting (LIM). We would have loved to meet you all in person, but the pandemic necessitated again this year an online meeting. Nonetheless, we have committed ourselves to provide you with the best possible online experience. We hope that you will find the meeting interesting, informative, and fun. Please remember that the short courses and all talks are being recorded. You will have access to all material until Spring 2022.

LIM's mandate is to bring together experts from a particular imaging-related topic every year. This year's theme—centered around Imaging for Deep Learning—is aimed to promote interaction between the imaging science and deep learning communities.

We have an exciting program ahead of us. On the first day we have scheduled two relevant short courses, one on imaging quality for automotive applications and the other on camera calibration for AR/VR, automotive, and machine vision applications. The following two days are devoted to the technical program and include two keynote talks, one invited lecture, five focal talks, eleven oral presentations, and six interactive poster presentations.

The Technical Program includes state-of-the-art papers in the field, with various topics including deep learning for image quality predictions, system characterization and optimization using deep learning, datasets for deep learning, deep neural networks for color applications, and imaging performance for deep neural networks.

During the interactive posters you can 'meet' and chat with posters authors about their research. We encourage you to join the virtual Posters Room, and the virtual Social Room where during session breaks you get the chance to meet with colleagues and network.

I'd like to take a moment to highlight our two keynote speakers: Joyce Farrell, Stanford University, shows how using simulated camera images, new metrics can be developed and tested for quantifying the effect that different camera parameters have on CNN performance, while Robin Jenkin, NVIDIA, explores desirable characteristics of image quality metrics, approaches, and the pitfalls of combining them, as well as strategies for ranking camera performance for use with autonomous systems. I'd also like to draw attention to the invited lecture Using Imaging Data for Efficient Colour Design to be delivered by Stephen Westland, University of Leeds.

Finally, I want to thank all the people who have helped to make this conference possible. These include Graham Finlayson, University of East Anglia, the LIM Series Chair; the Steering Committee composed of Michael Brown, York University, Susan Farnand, RIT, Rafal Mantiuk, Cambridge University, and Javier Vazquez-Corral, Universitat Autònoma de Barcelona; all the great colleagues and researchers who volunteered to serve on the Program Committee and as reviewers; and the staff at IS&T. I also want to thank the Institute of Physics who is in charge of running the online meeting.

I hope you enjoy your time at LIM2021!

—Sophie Triantaphillidou, LIM 2021 General Chair

CONFERENCE SPONSORS

IOP Institute of Physics
Printing and Graphics Science Group

THE RPS ROYAL PHOTOGRAPHIC SOCIETY
IMAGING SCIENCE GROUP

PROGRAM COMMITTEE

General Chair
Sophie Triantaphillidou
University of Westminster (UK)

Program Committee Members
Seyed Ali Amirshahi*
NTNU (Norway)
Simone Bianco*
University of Milano-Bicocca (Italy)
Sebastian Bosse*
Fraunhofer HHI (Germany)
Alexander Braun*
University of Applied Sciences Düsseldorf (Germany)
Peter Burns*
Burns Digital Imaging (US)
Christy Fernandez-Cull
Lyft / MIT (US)
Valentina Donzella*
University of Warwick (UK)
Chaker Larabi*
Universit of Poitiers (France)
Yi-Tun Lin*
University of East Anglia (UK)
Michal Mackiewicz
University of East Anglia (UK)
Ronald Müller
Vision Markets GmbH (Germany)
Marius Pedersen*
NTNU (Norway)
Jonathan Phillips*
consultant (US)
Alexandra Psarrou*
University of Westminster (UK)
Raquel Gil Rodríguez*
University of Giessen (Germany)
Raimondo Schettini*
University of Milano-Bicocca (Italy)
Giorgio Trumpy
University of Zurich (Switzerland)
Maria Vanrell*
Universitat Autònoma de Barcelona (Spain)
Barbara Villarini*
University of Westminster (UK)
Stephen Westland*
Leeds University (UK)

Steering Committee
Michael Brown*
York University (Canada)
Javier Vázquez-Corral*
Universitat Pompeu Fabra (Spain)
Susan Farnand
Rochester Institute of Technology (US)
Graham Finlayson*, series chair
University of East Anglia (UK)
Rafal Mantiuk*
University of Cambridge (UK)

*signifies paper reviewer

LIM 2021: SCHEDULE AND TECHNICAL PAPERS

MONDAY 20 SEPTEMBER 2021

SHORT COURSE PROGRAM

13:30 – 15:30 LONDON / 08:00 – 10:30 NY /
20:00 – 22:30 BEIJING

SC01: Imaging Quality for Automotive and Machine Vision Applications

Instructors: Don Williams, Image Science Associates, and Peter Burns, Burns Digital Imaging (US)

16:00 – 18:30 LONDON / 11:00 – 13:30 NY /
23:00 – 01:30 (TUES) BEIJING

SC02: Optics and Hardware Calibration of Compact Camera Modules for AR/VR, Automotive, and Machine Vision Applications

Instructors: Uwe Artmann, Image Engineering GmbH & Co. KG (Germany), and Kevin J. Matherson, Microsoft Corporation (US)

TUESDAY 21 SEPTEMBER 2021

WELCOME

Session Chairs: Graham Finlayson, University of East Anglia and Sophie Triantaphillidou, University of Westminster (UK)
09:30 – 09:40 LONDON / 04:30 – 04:40 NY /
16:30 – 16:40 BEIJING

SESSION I: DEEP LEARNING FOR IMAGE QUALITY AND AESTHETICS PREDICTIONS

Session Chair: Seyed Ali Amirshahi, Norwegian University of Science and Technology (Norway)
09:40 – 10:50 LONDON / 04:40 – 05:50 NY /
16:40 – 17:50 BEIJING

09:40 London / 04:40 NY / 16:40 Beijing

FOCAL TALK Deep Learning in Image Quality Assessment: Past, Present, and What Lies Ahead, *Seyed Ali Amirshahi, Norwegian University of Science and Technology (Norway).* **1**

Quality assessment of images plays an important role in different applications in image processing and computer vision. While subjective quality assessment of images is the most accurate approach due to issues objective quality metrics have been the go to approach. Until recently most such metrics have taken advantage of different handcrafted features. Similar (but with a slower speed) to other applications in image processing and computer vision, different machine learning techniques, more specifically Convolutional Neural Networks (CNNs) have been introduced in different tasks related to image quality assessment. In this short paper which is a supplement to a focal talk given with the same title at the London Imaging Meeting (LIM) 2021 we aim to provide a short timeline on how CNNs have been used in the field of image quality assessment so far, how the field could take advantage of CNNs to evaluate the image quality, and what we expect will happen in the near future.

10:10 London / 05:10 NY / 17:10 Beijing

Portrait Quality Assessment using Multi-scale CNN, *Nicolas Chahine and Salim Belkarfa, DXOMARK Image Labs (France)* **5**

In this paper, we propose a novel and standardized approach to the problem of camera – quality assessment on portrait scenes. Our goal is to evaluate the capacity of smartphone front cameras to preserve texture details on faces. We introduce a new portrait setup and an automated texture measurement. The setup includes two custom-built lifelike mannequin heads, shot in a controlled lab environment. The automated texture measurement includes a region-of-interest (ROI) detection and a deep neural network. To this aim, we create a realistic mannequins database, which contains images from different cameras, shot in several lighting conditions. The ground-truth is based on a novel pair-wise comparison technology where the scores are generated in terms of just-noticeable-differences (JND). In terms of methodology, we propose a multi-Scale CNN architecture with random crop augmentation, to overcome overfitting and to get a low-level feature extraction. We validate our approach by comparing its performance with several baselines inspired by the Image quality assessment (IQA) literature.

10:30 London / 05:30 NY / 17:30 Beijing

Modeling Image Aesthetics through Aesthetics-related Attributes, *Marco Leonardi[1], Paolo Napoletano[1], Alessandro Rozza[2], and Raimondo Schettini[1]; [1]University of Milano-Bicocca (Italy) and [2]lastminute.com (Switzerland)* **11**

Automatic assessment of image aesthetics is a challenging task for the computer vision community that has a wide range of applications. The most promising state-of-the-art approaches are based on deep learning methods that jointly predict aesthetics-related attributes and aesthetics score. In this article, we propose a method that learns the aesthetics score on the basis of the prediction of aesthetics-related attributes. To this end, we extract a multi-level spatially pooled (MLSP) features set from a pretrained ImageNet network and then these features are used to train a Multi Layer Perceptron (MLP) to predict image aesthetics-related attributes. A Support Vector Regression machine (SVR) is finally used to estimate the image aesthetics score starting from the aesthetics-related attributes. Experimental results on the "Aesthetics with Attributes Database" (AADB) demonstrate the effectiveness of our approach that outperforms the state of the art of about 5.5% in terms of Spearman's Rank-order Correlation Coefficient (SROCC).

SESSION BREAK

10:50 – 11:10 LONDON / 05:50 – 06:10 NY /
17:50 – 18:10 BEIJING

SESSION II: DATASETS FOR DEEP LEARNING

Session Chair: Jonas Unger, Linkoping University (Sweden)
11:10 – 12:20 LONDON / 06:10 – 07:20 NY /
18:10 – 19:20 BEIJING

11:10 London / 06:10 NY / 18:10 Beijing

FOCAL TALK Generative Inter-class Transformations for Imbalanced Data Weather Classification, *Apostolia Tsirikoglou, Marcus Gladh, Daniel Sahlin, Gabriel Eilertsen, and Jonas Unger, Linkoping University (Sweden)* **16**

This paper presents an evaluation of how data augmentation and inter-class transformations can be used to synthesize training data in low-data scenarios for single-image weather classification. In such scenarios, augmentations is a critical component, but there is a limit to how much improvements can be gained using classical augmentation strategies. Generative adversarial networks (GAN) have been demonstrated to generate impressive results, and have also been successful as a tool for data augmentation, but mostly for images of limited diversity, such as in medical applications. We investigate the possibilities in using generative augmentations for balancing a small weather classification dataset, where one class has a reduced number of images. We compare intra-class augmentations by means of classical transformations as well as noise-to-image GANs, to interclass augmentations where images from another class are transformed to the underrepresented class. The results show that it is possible to take advantage of GANs for inter-class augmentations to balance a small dataset for weather classification. This opens up for future work on GAN-based augmentations in scenarios where data is both diverse and scarce.

11:40 London / 06:40 NY / 18:40 Beijing

Visual Scan-path based Data-augmentation for CNN-based 360-degree Image Quality Assessment, *Abderrezzaq Sendjasni[1,2], Mohamed-Chaker Larabi[1], and Faouzi Alaya Cheikh[2]; [1]Université de Poitiers (France) and [2]Norwegian University of Science and Technology (Norway)* **21**

360-degree Image quality assessment (IQA) is facing the major challenge of lack of ground-truth databases. This problem is accentuated for deep learning based approaches where the performances are as good as the available data. In this context, only two databases are used to train and validate deep learning-based IQA models. To compensate this lack, a data-augmentation technique is investigated in this paper. We use visual scan-path to increase the learning examples from existing training data. Multiple scan-paths are predicted to account for the diversity of human observers. These scan-paths are then used to select viewports from the spherical representation. The results of the data-augmentation training scheme showed an improvement over not using it. We also try to answer the question of using the MOS obtained for the 360-degree image as the quality anchor for the whole set of extracted viewports in comparison to 2D blind quality metrics. The comparison showed the superiority of using the MOS when adopting a patch-based learning.

12:00 London / 07:00 NY / 19:00 Beijing

JIST-first: HDR4CV: High Dynamic Range Dataset with Adversarial Illumination for Testing Computer Vision Methods, *Param Hanji[1], Muhammad Z. Alam[1], Nicola Giuliani[2], Hu Chen[2], and Rafal K. Mantiuk[1]; [1]University of Cambridge (UK) and [2]Huawei Technologies (Germany)* **27**

Benchmark datasets used for testing computer vision (CV) methods often contain little variation in illumination. The methods that perform well on these datasets have been observed to fail under challenging illumination conditions encountered in the real world, in particular, when the dynamic range of a scene is high. The authors present a new dataset for evaluating CV methods in challenging illumination conditions such as low light, high dynamic range, and glare. The main feature of the dataset is that each scene has been captured in all the adversarial illuminations. Moreover, each scene includes an additional reference condition with uniform illumination, which can be used to automatically generate labels for the tested CV methods. We demonstrate the usefulness of the dataset in a preliminary study by evaluating the performance of popular face detection, optical flow, and object detection methods under adversarial illumination conditions. We further assess whether the performance of these applications can be improved if a different transfer function is used.

SESSION BREAK
12:20 – 13:20 LONDON / 07:20 – 08:20 NY / 19:20 – 20:20 BEIJING

TWO-MINUTE INTERACTIVE PAPER (POSTER) PREVIEWS AND INTERACTIVE SESSION

Session Chair: Barbara Villarini, University of Westminster (UK)
13:20 – 14:30 LONDON / 08:20 – 09:30 NY /
20:20 – 21:30 BEIJING
Interactive (Poster) Paper authors provide a brief overview of their work, followed by talks with authors in Gatherly. Posters may also be viewed in the LIM conference portal at any time.

Joint Unsupervised Infrared-RGB Video Registration and Fusion, *Imad Eddine Marouf[1,2], Luca Barras[2], Hakki Can Karaimer[2,3], Sabine Süsstrunk[2]; [1]Institut Polytechnique de Paris (France), [2]Ecole Polytechnique Federale de Lausanne (Switzerland), and [3]Advanced Micro Devices, Inc. (Canada)* **38**

We present a system to perform joint registration and fusion for RGB and Infrared (IR) video pairs. While RGB is related to human perception, IR is associated with heat. However, IR images often lack contour and texture information. The goal with the fusion of the visible and IR images is to obtain more information from them. This requires two completely matched images. However, classical methods assuming ideal imaging conditions fail to achieve satisfactory performance in actual cases. From the data-dependent modeling point of view, labeling the dataset is costly and impractical.

In this context, we present a framework that tackles two challenging tasks. First, a video registration procedure that aims to align IR and RGB videos. Second, a fusion method brings all the essential information from the two video modalities to a single video. We evaluate our approach on a challenging dataset of RGB and IR video pairs collected for fire-fighters to handle their tasks effectively in challenging visibility conditions such as heavy smoke after a fire.

Exposure problems, due to standard camera sensor limitations, often lead to image quality degradations such as loss of details and change in color appearance. The quality degradations further hiders the performances of imaging and computer vision applications. Therefore, the reconstruction and enhancement of under- and over-exposed images is essential for various applications. Accordingly, an increasing number of conventional and deep learning reconstruction approaches have been introduced in recent years. Most conventional methods follow color imaging pipeline, which strongly emphasize on the reconstructed color and content accuracy. The deep learning (DL) approaches have conversely shown stronger capability on recovering lost details. However, the design of most DL architectures and objective functions don't take color fidelity into consideration and, hence, the analysis of existing DL methods with respect to color and content fidelity will be pertinent. Accordingly, this work presents performance evaluation and results of recent DL based overexposure reconstruction solutions. For the evaluation, various datasets from related research domains were merged and two generative adversarial networks (GAN) based models were additionally adopted for tone mapping application scenario. Overall results show various limitations, mainly for severely over-exposed contents, and a promising potential for DL approaches, GAN, to reconstruct details and appearance.

Blind assessment of video quality is a widely covered topic in computer vision. In this work, we perform an analysis of how much the effectiveness of some of the current No-Reference VQA (NR-VQA) methods varies with respect to specific types of scenes. To this end, we automatically annotated the videos from two video quality datasets with user-generated videos whose content is unknown and then estimated the correlation for the different categories of scenes. The results of the analysis highlight that the prediction errors are not equally distributed among the different categories of scenes and indirectly suggest what next generation NR-VQA methods should take into account and model.

We demonstrate that a deep neural network can achieve near-perfect colour correction for the RGB signals from the sensors in a camera under a wide range of daylight illumination spectra. The network employs a fourth input signal representing the correlated colour temperature of the illumination. The network was trained entirely on synthetic spectra and applied to a set of RGB images derived from a hyperspectral image dataset under a range of daylight illumination with CCT from 2500K to 12500K. It produced an invariant output image as XYZ referenced to D65, with a mean colour error of approximately *1.0* ΔE^*_{ab}.

We present a novel metric Spatial Recall Index to assess the performance of machine-learning (ML) algorithms for automotive applications, focusing on where in the image which performance occurs. Typical metrics like intersection-over-union (IoU), precision-recall-curves or average precision (AP) quantify the performance over a whole database of images, neglecting spatial performance variations. But as the optics of camera systems are spatially variable over the field of view, the performance of ML-based algorithms is also a function of space, which we show in simulation: A realistic objective lens based on a Cooke-triplet that exhibits typical optical aberrations like astigmatism and chromatic aberration, all variable over field, is modeled. The model is then applied to a subset of the BDD100k dataset with spatially-varying kernels. We then quantify local changes in the performance of the pre-trained Mask R-CNN algorithm. Our examples demonstrate the spatial dependence of the performance of ML-based algorithms from the optical quality over field, highlighting the need to take the spatial dimension into account when training ML-based algorithms, especially when looking forward to autonomous driving applications.

(Work in Progress) Surgical simulation can provide a useful way to train for the complex skills required by surgery in a safe environment. However, it requires enough realism for the surgeons to be engaged in a challenging way. This research aims to develop a hybrid surgical simulator for laparoscopic surgery on the bile duct, combining a silicone-based physical model, and an image processing technique to improve the visual realism of the simulation. The image processing technique is style transfer, a technique where a CNN is trained to apply the style from a chosen image onto other images. In the context of surgical simulation, it can apply the style from a picture of surgery on the video frames taken by the endoscopic camera during the simulation, resulting in a more realistic simulation practice.

SESSION III: COLOR AND CONSTANCY

Session Chair: Simone Bianco, University of Milano – Bicocca (Italy)
14:30 – 15:40 LONDON / 09:30 – 10:40 NY /
21:30 – 22:40 BEIJING

14:30 London / 09:30 NY / 21:30 Beijing

In this article we show the change in paradigm occurred in color constancy algorithms: from a pre-processing step in image understanding, to the exploitation of image understanding and computer vision results and techniques. Since color constancy is an ill-posed problem, we give an overview of the assumptions on which classical color constancy algorithms are based in order to solve it. Then, we chronologically review the color constancy algorithms that exploit results and techniques borrowed from the image understanding research field in order to exploit assumptions that could be met in a larger number of images.

15:00 London / 10:00 NY / 22:00 Beijing

Revisiting and Optimising a CNN Colour Constancy Method for Multi-illuminant Estimation, *Ghalia Hemrit[1,2] and Joseph Meehan[2]; [1]Huawei Technologies (France) and [2]University of East Anglia (UK).* **68**

The aim of colour constancy is to discount the effect of the scene illumination from the image colours and restore the colours of the objects as captured under a 'white' illuminant. For the majority of colour constancy methods, the first step is to estimate the scene illuminant colour. Generally, it is assumed that the illumination is uniform in the scene. However, real world scenes have multiple illuminants, like sunlight and spot lights all together in one scene. We present in this paper a simple yet very effective framework using a deep CNN-based method to estimate and use multiple illuminants for colour constancy. Our approach works well in both the multi and single illuminant cases. The output of the CNN method is a region-wise estimate map of the scene which is smoothed and divided out from the image to perform colour constancy. The method that we propose outperforms other recent and state of the art methods and has promising visual results.

15:20 London / 10:20 NY / 22:20 Beijing

CMYK-CIELAB Color Space Transformation using Machine Learning Techniques, *Ronny Velastegui Sandoval and Marius Pedersen, Norwegian University of Science and Technology (Norway)* **73**

In this work four different machine learning approaches have been implemented to perform the color space transformation between CMYK and CIELAB color spaces. We have explored the performance of Support-Vector Regression (SVR), Artificial Neural Networks (ANN), Deep Neural Networks (DNN), and Radial Basis Function (RBF) models to achieve this color space transformation, both AToB and BToA direction. The data set used for this work was FOGRA53 which is composed of 1617 color samples represented both in CMYK and CIELAB color space values. The accuracy of the transformation models was measured in terms of ΔE* color difference. Moreover, the proposed models were compared, in practical terms, with the performance of the standard ICC profile for this color space transformation. The results showed that, for the forward transformation (CMYK to CIELAB), the highest accuracy was obtained using RBF. While, for the backward transformation (CIELAB to CMYK), the highest accuracy was obtained with DNN.

SESSION BREAK

15:40 – 16:00 LONDON / 10:40 – 11:00 NY /
22:40 – 23:00 BEIJING

CONFERENCE KEYNOTE I

Session Chair: Sophie Triantaphillidou, University of Westminster (UK)
16:00 – 17:00 LONDON / 11:00 – 12:00 NY /
23:00 – 24:00 (WED) BEIJING

Soft-prototyping Camera Designs for Autonomous Driving, *Joyce E. Farrell, Stanford University Center for Image Systems Engineering, SCIEN (US)* . . . . . *

It is impractical to build different cameras and then acquire and label the necessary data for every potential camera design. Creating software simulations that can generate synthetic camera images captured in physically realistic 3D scenes (soft prototyping) is the only practical approach. We implemented soft-prototyping tools that can quantitatively simulate image radiance and camera designs to create synthetic camera images that are input to convolutional neural networks for car detection. We show that performance derived from training on physically-based multispectral simulations of camera images generalizes to real camera images with nearly the same performance level as training based on real camera image datasets. Using simulations, we can develop and test new metrics for quantifying the effect that different camera parameters have on CNN performance. As an example, we introduce a new metric based on the distance at which object detection reaches 50%. Our open-source and freely available prototyping tools, together with performance-based metrics, enable us to evaluate the effect that changes in scene and camera parameters have on CNN performance.

WEDNESDAY 22 SEPTEMBER 2021

WELCOME

Session Chairs: Graham Finlayson, University of East Anglia and Sophie Triantaphillidou, University of Westminster (UK)
10:30 – 10:40 LONDON / 05:30 – 05:40 NEW YORK /
17:30 – 17:40 BEIJING

SESSION IV: IMAGING PERFORMANCE FOR DEEP LEARNING

Session Chair: Valentina Donzella, University of Warwick (UK)
10:40 – 11:50 LONDON / 05:40 – 06:50 NY /
17:40 – 18:50 BEIJING

10:40 London / 05:40 NY / 17:40 Beijing

FOCAL TALK The Data Conundrum: Compression of Automotive Imaging Data and Deep Neural Network based Perception, *Pak Hung Chan[1], Georgina Souvalioti[1], Anthony Huggett[2], Graham Kirsch[2], and Valentina Donzella[1]; [1]University of Warwick and [2]ON Semiconductors (UK).* **78**

Video compression in automated vehicles and advanced driving assistance systems is of utmost importance to deal with the challenge of transmitting and processing the vast amount of video data generated per second by the sensor suite which is needed to support robust situational awareness. The objective of this paper is to demonstrate that video compression can be optimised based on the perception system that will utilise the data. We have considered the deployment of deep neural networks to implement object (i.e. vehicle) detection based on compressed video camera data extracted from the KITTI MoSeg dataset. Preliminary results indicate that re-training the neural network with M-JPEG compressed

**Abstract only; no proceedings paper.*

videos can improve the detection performance with compressed and uncompressed transmitted data, improving recalls and precision by up to 4% with respect to re-training with uncompressed data.

11:10 London / 06:10 NY / 18:10 Beijing

Impact of the Windshield Optical Aberrations on Visual Range Camera based Classification Tasks Performed by CNNs, *Christian Krebs, AGP Europe GmbH; and Patrick Müller and Alexander Braun, Hochschule Düsseldorf (Germany)* **83**

Cameras operating in the visual range of the electromagnetic spectrum are central to advanced driver assistance systems (ADAS). Front cameras, analyzing traffic, are often located behind the windshield to detect and classify objects. Thus, the area of the windshield within the camera's field of view is a part of the optical system. Simple windshields consist of two curved glass surfaces connected by a thermoplastic interlayer. Due to defects present in the raw glass, as well as those introduced during the bending and lamination process, windshields will have optical aberrations. While optical quality may be suitable for human vision, it can fall short of what is needed for machine vision. In this article we investigate how the optical aberrations generated by laminated safety glass (LSG) influence the optical performance of a camera system and based on this, how the classification of image content by a convolutional neural network (CNN) is affected. A method for wavefront measurements of LSG samples is presented, which allows us to parameterize a linear optical model in Zernike Space. From this, we derive space-variant point spread functions (PSFs) and apply those to the dataset to simulate the windshield's impact on the camera image. As a use case, a CNN was trained on the unmodified dataset and compared to the modified versions with the LSG models applied. We measured and modelled two different LSG samples, one with high and the other one with low optical quality. We compare the prediction accuracy of the classification with the unmodified data. The high-quality sample had negligible effect on the overall classification accuracy, while the low-quality sample lowered the prediction accuracy by up to ten percentage points due to the optical aberrations.

11:30 London / 06:30 NY / 18:30 Beijing

Natural Scene Derived Camera Edge Spatial Frequency Response for Autonomous Vision Systems, *Oliver van Zwanenberg, Sophie Triantaphillidou, Robin Jenkin, and Alexandra Psarrou, University of Westminster (UK)* **88**

The edge-based Spatial Frequency Response (e-SFR) is an established measure for camera system quality performance, traditionally measured under laboratory conditions. With the increasing use of Deep Neural Networks (DNNs) in autonomous vision systems, the input signal quality becomes crucial for optimal operation. This paper proposes a method to estimate the system e-SFR from pictorial natural scene derived SFRs (NSSFRs) as previously presented, laying the foundation for adapting the traditional method to a real-time measure.

In this study, the NS-SFR input parameter variations are first investigated to establish suitable ranges that give a stable estimate. Using the NS-SFR framework with the established parameter ranges, the system e-SFR, as per ISO 12233, is estimated. Initial validation of results is obtained from implementing the measuring framework with images from a linear and a non-linear camera system. For the linear system, results closely approximate the ISO 12233 e-SFR measurement. Non-linear system measurements exhibit scene-dependant characteristics expected from edge-based methods. The requirements to implement this method in real-time for autonomous systems are then discussed.

SESSION BREAK
11:50 – 13:00 LONDON / 06:50 – 08:00 NY /
18:50 – 20:00 BEIJING

INVITED LECTURE

Session Chair: Sophie Triantaphillidou, University of Westminster (UK)
13:00 – 14:00 LONDON / 08:00 – 09:00 NY /
20:00 – 21:00 BEIJING

Using Imaging Data for Efficient Colour Design, *Stephen Westland, University of Leeds and Colour Intelligence Ltd. (UK)* . . . . . *

Smart manufacturing promises to radically change how products are manufactured and this could lead to significant improvements in efficiency and sustainability. Machine learning is a core technology that underpins the potential of smart manufacturing to bring about this required change. This presentation will focus on an unsupervised machine learning clustering technique and its application to colour image data. Recent work to automate extraction of colour palettes from images will be described including application of this method to use with mood boards, fashion shows, and to explore colour meaning. A method for predicting the visual similarity of colour palettes will also be introduced. Some novel extensions of clustering will be described and the combination of the unsupervised clustered method with more complex supervision techniques (e.g., deep learning) will be demonstrated. Potential applications can enable "designing for one" or personalisation to become one step closer.

SESSION BREAK
14:00 – 14:20 LONDON / 09:00 – 09:20 NY /
21:00 – 21:20 BEIJING

CONFERENCE KEYNOTE II

Session Chair: Sophie Triantaphillidou, University of Westminster (UK)
14:20 – 15:20 LONDON / 09:20 – 10:20 NY /
21:20 – 22:20 BEIJING

Camera Metrics for Autonomous Vision, *Robin Jenkin, NVIDIA (US) and University of Westminster (UK)* . . . . . *

Not all pixels are created equal, neither are all lenses, or sensors, or manufacturers. This causes a large variance in image quality from cameras with nominally the same fundamental specifications, such as pixel size, focal length, and f-number. Individual objective camera metrics can provide insight into the sharpness or noise performance of cameras, for example, and instinctively we desire more of everything. This, however, represents unconstrained development. Because of module size constraints, we are often in the position of not being able to arbitrarily increase pixel size without reducing pixel count, for example. Or the budget for an extra two surfaces in our lens may not exist. Or equally for the more expensive higher-yielding manufacturing process.

In the above circumstances we are forced to trade performance in one image quality dimension for another. To do this we need to understand the contribution to overall camera performance from individual quality measures in order to trade them fairly. One unit of sharpness may not be equal to one unit of noise. Further combined with complications due to environmental conditions, such as illumination levels, motion, target, and background complexity, it is difficult to distinguish between performance limitations imposed by the imaging system and those of the DNN itself?

Unfortunately, far too often in papers exploring DNN performance,

**Abstract only; no proceedings paper.*

the description of the images used is limited to the pixel count, total number, and split between training and validation sets.

This talk explores some desirable characteristics of image quality metrics, approaches, and pitfalls of combining them and some strategies for ranking camera performance for use with autonomous systems.

SESSION BREAK
15:20 – 15:40 LONDON / 10:20 – 10:40 NY /
22:20 – 22:40 BEIJING

SESSION V: DEEP LEARNING FOR CHARACTERIZATION AND OPTIMIZATION

Session Chair: Ray Ptucha, Apple Inc. (US)
15:40 – 17:10 LONDON / 10:40 – 12:10 NY /
22:40 – 00:10 (THURS) BEIJING

15:40 London / 10:40 NY / 18:20 Beijing
Towards a Generic Neural Network Architecture for Approximating Tone Mapping Algorithms, *Jake McVey and Graham Finlayson, University of East Anglia (UK)* **93**
Tone curves are a key feature in any image processing pipeline, and are used to change the pixel values of an input image to find an output image that looks better. Perhaps the most widely deployed tone curve algorithm is Contrast Limited Histogram Equalisation (CLHE). CLHE is an iterative algorithm that tone maps an input image so that the histogram of the output is (approximately) maximally uniform subject to the constraint that the tone curve has bounded slope (neither too large or too small).

In this paper, we build upon a neural network framework that was recently developed to deliver CLHE in fewer iterations (each layer in the neural network is analogous to a single iteration of CLHE, but the network has fewer layers than the number of iterations needed by CLHE). The key contribution of this paper is to show that the same network architecture can be used to implement a more complex (and more powerful) tone mapping algorithm. Experiments validate our method.

16:00 London / 11:00 NY / 23:00 Beijing
A Study of Neural Network-based LCD Display Characterization, *Joan Prats-Climent[1], Luis Gómez-Robledo[2], Rafael Huertas[2], Sergio García-Nieto[1], María José Rodríguez-Álvarez[1], and Samuel Morillas[1]; [1]Universitat Politécnica de Valéncia and [2]Univeridad de Granada (Spain).* **97**
In this paper we study up to what extent neural networks can be used to accurately characterize LCD displays. Using a programmable colorimeter we have taken extensive measures for a DELL Ultrasharp UP2516D to define training and testing data sets that are used, in turn, to train and validate two neural networks: one of them using tristimulus values, XYZ, as inputs and the other one color coordinates, xyY . Both networks have the same layer structure which has been experimentally determined. The errors from both models, in terms of DE00 color difference, are analysed from a colorimetric point of view and interpreted in order to understand how both networks have learned and how is their performance in comparison with other classical models. As we will see, the comparison is in average in favor of the proposed models but it is not better in all cases and regions of the color space.

16:20 London / 11:20 NY / 23:20 Beijing
Automatic Noise Analysis on Still Life Chart, *Salim Belkarfa, Ahmed Hakim Choukarah, and Marcelin Tworski, DXOMARK Image Labs (France)* **101**
In this paper, we tackle the issue of estimating the noise level of a camera, on its processed still images and as perceived by the user. Commonly, the characterization of the noise level of a camera is done using objective metrics determined on charts containing uniform patches at a given condition. These methods can lead to inadequate characterizations of the noise of a camera because cameras often incorporate denoising algorithms that are more efficient on uniform areas than on areas containing details. Therefore, in this paper, we propose a method to estimate the perceived noise level on natural areas of a still-life chart. Our method is based on a deep convolutional network trained with ground truth quality scores provided by expert annotators. Our experimental evaluation shows that our approach strongly matches human evaluations.

16:40 London / 11:40 NY / 23:40 Beijing
FOCAL TALK Mitigating Limitations of Deep Neural Networks for Imaging Systems, *Ray Ptucha, Apple Inc. (US)* . . . . . *
Deep learning has been shown to advantage many aspects of traditional imaging and machine vision systems. Its applications are ubiquitous, from security, medical, transportation, communication and entertainment. Convolutional neural networks (CNNs) are end-to-end, learning low level features and classifier simultaneously in a supervised, semi-supervised, or self-supervised fashion, giving substantial advantage over methods using independently solved features and classifiers. Although modern deep networks have substantially improved imaging applications, they 1) are often forced to make a classification decision even when the input is noise; 2) may have >100M weights, needing lots of training data; and 3) many problems cannot be factored into nice gridded structures required by most neural engines. This talk will firstly focus on understanding these problem spaces, and then borrow mitigation techniques from other disciplines to further the imaging landscape.

BEST PAPER AWARD PRESENTATION AND CLOSING REMARKS

Session Chairs: Graham Finlayson, University of East Anglia, and Sophie Triantaphillidou, University of Westminster (UK)
17:10 – 17:30 LONDON / 12:10 – 12:30 NY /
00:10 – 00:30 (THURS) BEIJING

APPENDIX: WORK-IN-PROGRESS ABSTRACT

Can style transfer improve the realism of simulation of laparoscopic bile duct exploration using ultrasound?

Marine Shao and David Huson; University of the West of England; Bristol, UK

Abstract

Surgical simulation has demonstrated its potential to improve surgeons' performance through practice [1]. However, surgeons need to be engaged in a challenging way which can only be achieved through a realistic practice. There are several types of surgical simulators, including physical simulators, virtual-reality-based simulators, and hybrid simulators [2]. Physical simulators can provide tactile feedback, but are not very realistic-looking and can only provide a limited number of scenarios and anatomies. Virtual-reality presents more diversity of scenarios and pathologies but the tactile feedback is limited. Hybrid simulations aims to combine both methods to provide better training.

The aim of this research project is to develop a hybrid simulator for laparoscopic bile duct exploration using ultrasound. To our knowledge, there is no satisfying simulator able to teach the challenging steps of this surgery, such as suturing of the bile duct and intra-operative ultrasounds. The simulator needs to feel realistic when the surgeons suture the bile duct and to be ultrasound-visible to allow radiographic examination.

This work focuses on developing a simulator using ultrasound visible materials. Most ultrasound simulators are gel-based resulting in a low-shelf-life, or are made of silicones which have a longer shelf-life but do not look as realistic [3]; ; however, during a laparoscopy, the surgeon can only observe the surgery through a camera, for this reason, it is possible to enhance the visual realism through images processing of the video.

The utilisation of an image-processing technique, known as style-transfer, is investigated for its potential to improve the realism of a silicone ultrasound simulator. Style transfer aims to combine a content image and a style image, to produce an output image combining the features of the content image and the style of the style image [4], [5]. Previous researches also demonstrate the potential of image processing techniques to enhance surgical simulators [6,7], using style transfer or image-to-image translation. In this work, it is tested to enhance the realism of the ultrasound images and the endoscopic images taken during the simulation practice.

Results shows that this method can implement the style of real images from surgery onto a silicone-based model to improve its appearance. The algorithm works in real-time, allowing its use in a simulation practice.

One of the limitations of the simulator is that the speed of sound in the silicone is around 1000m/s instead of 1540m/s in the tissues; as a result, the ultrasound images of the simulator are deformed, requiring further image processing to correct the dimensions. Furthermore, the generalization of this technique can be limited by the technology used to perform the ultrasound examination. Further work would include an evaluation of the outcome by surgeons. [5]

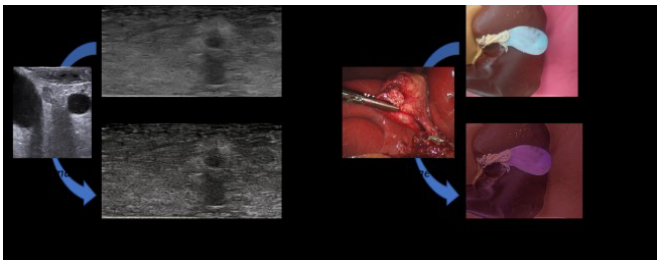

Figure 1. *Results of the style transfer on an ultrasound simulation (a) and on an endoscopic image (b)*

References

[1] D. R. Stather, P. MacEachern, K. Rimmer, C. A. Hergott, and A. Tremblay, "Assessment and learning curve evaluation of endobronchial ultrasound skills following simulation and clinical training," Respirology, vol. 16, no. 4, pp. 698–704, 2011, doi: 10.1111/j.1440-1843.2011.01961.x.

[2] C. Freschi, S. Parrini, N. Dinelli, M. Ferrari, and V. Ferrari, "Hybrid simulation using mixed reality for interventional ultrasound imaging training," Int. J. Comput. Assist. Radiol. Surg., vol. 10, no. 7, pp. 1109–1115, 2015, doi: 10.1007/s11548-014-1113-x.

[3] M. S. Ahmad et al., "Chemical Characteristics, Motivation and Strategies in choice of Materials used as Liver Phantom: A Literature Review," J. Med. Ultrasound, vol. 28, Jan. 2020, doi: 10.4103/JMU.JMU_4_19.

[4] L. Gatys, A. Ecker, and M. Bethge, "A Neural Algorithm of Artistic Style," J. Vis., vol. 16, no. 12, p. 326, 2016, doi: 10.1167/16.12.326.

[5] H. Huang et al., "Real-time neural style transfer for videos," Proc. - 30th IEEE Conf. Comput. Vis. Pattern Recognition, CVPR 2017, vol. 2017-Janua, pp. 7044–7052, 2017, doi: 10.1109/CVPR.2017.745.

[6] I. Luengo, E. Flouty, P. Giataganas, P. Wisanuvej, J. Nehme, and D. Stoyanov, "Surreal: Enhancing Surgical simulation Realism using style transfer," Br. Mach. Vis. Conf. 2018, BMVC 2018, pp. 1–12, 2019.

[7] S. Engelhardt, L. Sharan, M. Karck, R. De Simone, and I. Wolf, "Cross-Domain Conditional Generative Adversarial Networks for Stereoscopic Hyperrealism in Surgical Training," Inform. aktuell, p. 341, 2020, doi: 10.1007/978-3-658-29267-6_75.