Waqas Sultani

Contact Information

Email: waqas5163@gmail.com

About me

I am a US-trained Computer vision engineer (Ph.D.) with more than 12 years of extensive hands-on experience working on Computer Vision and Machine Learning. Specifically, I have been analyzing videos and images using supervised, weakly supervised, and unsupervised machine learning techniques. Along with academia, I have spent sufficient time in the industry. Recently, I spent one year in Hazenai working as a Principle Machine Learning Engineer. At Hazenai, I have mainly worked on designing and implementing improved multiple object tracking and robust traffic phase detection.

I have experience working on projects from many different companies and organizations including the US Intelligence Advanced Research Projects Activity (IARPA), US Data Transport Solution (DTS), US National Institute of Justice (NIJ), Xerox (USA), Facebook Research (US), Google Research (US) Samsung (South Korea), and Hazen.ai (Saudi Arabia).

As a research scientist, my recent research projects related to human action recognition, anomaly detection, small object detection, geolocalization, and medical imaging got published in the top venues including CVPR, ICRA, AAAI, WACV, CVIU, MICCAI, etc. I was awarded the Facebook-CV4GC research award in 2019 and a Google Research scholar in 2023 for projects related to medical imaging.

I did my MS from Seoul National University (the 27th best university in the world at that time) and Ph.D. (under Mubarak Shah) from the Center for Research in Computer Vision (6th best Computer vision research group in the US), University of Central Florida, USA.

Research Interest

Deep Learning, Computer Vision, Medical Imaging, Machine learning, Optimization Methods, Weakly supervised Detection and Classification

Education

Ph.D. Computer Science (Computer Vision), **2011-2017**

- Center for Research in Computer Vision, University of Central Florida, USA
- Advisor: Dr. Mubarak Shah, Trustee Chair Professor
- Dissertation: Weakly Labeled Action Recognition and Detection

MSc. Computer Engineering, **2008-2010**

- Seoul National University, South Korea
- Perceptional and Intelligence lab
- Advisor: Dr. Jin Young Choi

BSc. Electrical Engineering, **2003-2006**

- University of Engineering and Technology, Taxila, Pakistan
- Major: Electronics
- Advisor: Dr. Muhammad Amin

Positions

Assistant Professor **2018- Present**
• Information Technology University, Lahore

Principle Machine Learning Scientist **2022-2023**
• Hazen.ai

Graduate Research Assistant **2011-2017**
• University of Central Florida, USA

Summer Intern **2014**
• Xerox, New York, USA

Computer Vision Scientist **2010-2011**
• COMSATS University, Islamabad

Graduate Research Assistant **2008-2010**
• Seoul National University, South Korea

System Design Engineer **2007-2008**
• Advanced Engineering Organization

Awards, Grants

• **Google**, Research Scholar Award, "A Large-Scale Blood Cancer Dataset of Multiple Microscopes for Low-Cost Blood Cancer Detection", Role: PI

• **Facebook**, Computer Vision for Global Challenges research award, "Low-cost deep learning solution to real-time detection of malaria", Role: PI

• HEC, **Startup Research Grant** Program, "Criminal activities detection in CCTV video using computer vision techniques", Role: PI

• HEC, **Crime Investigation and Prevention Lab**, Role: Co-PI project funded by HEC

• **MS scholarship** by Higher Education Commission, Pakistan

• Recipient of **NSF-USA** Innovation-corps training

• **Travel Grants** for CVPR, 2014 and ICPR, 2010

• Islamabad College for Boys scholarship for the top position

Patents

• Robert P. Loce, **Waqas Sultani**, Beilei Xu, Hao Wu, "System and Method for Seat Occupancy Detection from Ceiling Mounted Camera using Robust Adaptive Threshold Criteria", **US Patent: US9378421 B2.**

• Robert P. Loce, **Waqas Sultani**, Hao Wu, Beilei Xu, Thomas F. Wade, Mary Ann Sprague, Patricia Swenton-Wall, Megan Clar, Eric Harte, "System and Method for Detecting Settle Down Time using Computer Vision Techniques", **US Patent: US9384396 B2.**

Selected Publications

• Tushar Sangam, Ishan Rajendrakumar Dave, **Waqas Sultani**, Mubarak Shah "Transvisdrone: Spatio-temporal transformer for vision-based drone-to-drone detection in aerial videos ", *IEEE International Conference on Robotics and Automation,* ***ICRA****, 2023*

• Xiaohan Zhang, Xingyu Li, **Waqas Sultani**, Yi Zhou, Safwan Wshah "Cross-view Geo-localization via Learning Disentangled Geometric Layout Correspondence ", *The Association for the Advancement of Artificial Intelligence,* ***AAAI****, 2023*

• Xiaohan Zhang, **Waqas Sultani**, Safwan Wshah "Cross-View Image Sequence Geo-localization ", *IEEE Winter Conference on Computer Vision,* ***WACV****, 2023*

• **Waqas Sultani**, Wajahat Nawaz, Syed Javed, Muhammad Sohail Danish, Asma Saadia, Mohsen Ali "Towards Low-Cost and Efficient Malaria Detection ", *IEEE Conference on Computer Vision and Pattern Recognition,* ***CVPR****, 2022*

• Muhammad Waseem Ashraf, **Waqas Sultani**, Mubarak Shah "Dogfight Detecting Drones from Drones Videos", *IEEE Conference on Computer Vision and Pattern Recognition,* ***CVPR****, 2021*

• Adnan Qayyum, **Waqas Sultani**, Fahad Shamshad, Junaid Qadir, and Rashid Tufail, "Single-Shot Retinal Image Enhancement Using Deep Image Priors", *International Conference on Medical Image Computing and Computer-Assisted Intervention,* ***MICCAI****, 2020*

• **Waqas Sultani**, Chen Chen, Mubarak Shah, "Real-world anomaly detection in surveillance videos", *IEEE Conference on Computer Vision and Pattern Recognition,* ***CVPR****, 2018*

• **Waqas Sultani**, Mubarak Shah, "What if we do not have multiple videos of the same action? - Video Action Localization Using Web Images", *IEEE Conference on Computer Vision and Pattern Recognition, **CVPR**, 2016*

• **Waqas Sultani**, Imran Saleemi, "Human Action Recognition across Datasets by Foreground Focused Histogram Decomposition", *IEEE Conference on Computer Vision and Pattern Recognition, **CVPR**, 2014*

• **Waqas Sultani**, Jin Young Choi, "Abnormal Traffic Detection using Intelligent Driver Model", *IEEE International Conference on Pattern Recognition, **ICPR**, 2010*

Journal Publications

• Adnan Qayyum, **Waqas Sultani**, Fahad Shamshad, Rashid Tufail, Junaid Qadir "Single-shot retinal image enhancement using untrained and pretrained neural networks priors integrated with analytical image priors", *Journal of Computers in Biology and Medicine, **CBM**, 2022*

• Muhammad Zaid, Shafaqat Ali, Mohsen Ali, Sarfaraz Hussein, Asma Saadia, **Waqas Sultani** "Identifying out of distribution samples for skin cancer and malaria images", *Journal of Biomedical Signal Processing and Control, **BSPC**, 2022*

• M Fasi Ur Rehman, Izza Aftab, **Waqas Sultani**, Mohsen Ali "Mapping Temporary Slums From Satellite Imagery Using a Semi-Supervised Approach", *IEEE Geoscience and Remote Sensing Letters, **GRSL**, 2022*

• Bilal Yousaf, Muhammad Usama, **Waqas Sultani**, Arif Mahmood Junaid Qadir "Fake visual content detection using two-stream convolutional neural networks", *Journal of Neural Computing and Applications volume, **NCA**, 2022*

• Qazi Ammar Arshad, Mohsen Ali, Saeed-ul Hassan, Chen Chen, Ayisha Imran, Ghulam Rasul, **Waqas Sultani** "A dataset and benchmark for malaria life-cycle classification in thin blood smear images", *Journal of Neural Computing and Applications volume, **NCA**, 2021*

• Nadeem Yousaf, Sarfaraz Hussein, **Waqas Sultani** "Estimation of BMI from facial images using semantic segmentation based region-aware pooling", *Journal of Computers in Biology and Medicine, **CBM**, 2021*

• **Waqas Sultani**, Mubarak Shah "Human Action Recognition in Drone Videos Using a Few Aerial Training Examples", *Journal of Computer Vision and Image Understanding, **CVIU**, 2021*

• Usman Ali, **Waqas Sultani**, Mohsen Ali, "Destruction detection from sky: Weakly supervised approach for destruction detection in satellite imagery", *Journal of Photogrammetry and Remote Sensing, **ISPRS**, 2020*

• Anza Shakeel, **Waqas Sultani**, Mohsen Ali, "Deep built-structure counting in satellite imagery using attention based re-weighting", *Journal of Photogrammetry and Remote Sensing, **ISPRS**, 2019*

• **Waqas Sultani**, Mubarak Shah, "Automatic Action Annotation in Weakly Labeled Videos", *Journal of Computer Vision and Image Understanding, **CVIU**, 2017*

• **Waqas Sultani**, Dong Zhang, Mubarak Shah, "Unsupervised Action Proposal Ranking through Proposal Recombination", *Journal of Computer Vision and Image Understanding, **CVIU**, 2017*

• **Waqas Sultani**, Soroush Mokhtari and Hae-Bum Yun, "Automatic Pavement Object Detection using Superpixel Segmentation Combined with Conditional Random Field", *IEEE Transactions on Intelligent Transportation Systems, **IEEE Trans ITS**, 2017*

Related Research Activities

Moving Small Object Detection in videos: Paper link, Code link, Project page, Slides link
As airborne vehicles are becoming more autonomous and ubiquitous, it has become vital to develop the capability to detect the objects in their surroundings. In this project, we attempt to address the problem of drones detection from other flying drones. The erratic movement of the source and target drones, small size, arbitrary shape, large intensity variations, and occlusion make this problem quite challenging. In this scenario, region-proposal-based methods are not able to capture sufficient discriminative foreground-background information.

Also, due to the extremely small size and complex motion of the source and target drones, feature aggregation-based methods are unable to perform well. To handle this, instead of using region-proposal-based methods, we propose to use a two-stage segmentation-based approach employing spatio-temporal attention cues. During the first stage, given the overlapping frame regions, detailed contextual information is captured over convolution feature maps using pyramid pooling. After that pixel and channel-wise attention are enforced on the feature maps to ensure accurate drone localization. In the second stage, first-stage detections are verified and new probable drone locations are explored. To discover new drone locations, motion boundaries are used. This is followed by tracking candidate drone detections for a few frames, cuboid formation, extraction of the 3D convolution feature map, and drone detection within each cuboid. The implements are mostly done in Keras.

Domain Adaptive Object Detection: Paper link
Similar to other supervised deep learning methods, object detection methods trained on the source domain do not generalize adequately to a new target domain. This problem, known as domain shift could be exhibited by the change in style, camera poses, object size and orientation, or the number or location of objects in the scene, among other things. This problem is also prominent in data captured by different medical instruments. We present a new dataset to further the research on malaria microscopy over low-cost microscopes at low magnification. Our large-scale dataset consists of images of blood-smear slides from several malaria-infected patients, collected through microscopes at two different cost spectrums and multiple magnifications. Malarial cells are annotated for the localization and life-stage classification task on the images collected through the high-cost microscope at high magnification. We design a mechanism to transfer these annotations from the high-cost microscope at high magnification to the low-cost microscope, at multiple magnifications. Multiple object detectors and domain adaptation methods are presented as baselines. Furthermore, a partially supervised domain adaptation method is introduced to adapt the object detector to work on the images collected from the low-cost microscope. The implementation is mostly done in Pytorch.

Object Segmentation:
(1) Supervised Object Segmentation: Paper link
This project is mainly focused on different pavement object segmentation. Pavement images contain various objects, such as lane-marker, manhole covers, patches, potholes, and curbing. Accurate and robust computer vision algorithms are necessary to detect these various objects that have random shapes, colors, and sizes. In this paper, we have addressed the problem of automatic object detection in pavement images using a unified framework. To detect an object of arbitrary shape efficiently, we first divide the image into small consistent regions called superpixels. We then compute several texture and intensity features within each superpixel. After that, we train a support vector machine (SVM) classifier for every feature separately in the one-verses-all paradigm. Since these superpixels' probabilistic scores are independently computed, they do not preserve neighborhood consistency. Therefore, to enforce superpixel neighborhood label consistency, we use contextual optimization techniques i.e., conditional random field (CRF).
(2) Weakly-Supervised Object Segmentation: Paper link, Code link
This project is focused on object segmentation in scenarios where it is difficult to obtain pixel-level annotations. Therefore, we employ image-level annotations to obtain pixel-wise segmentation. We model the destruction in satellite imagery using a deep learning model employing a weakly-supervised approach. In stark contrast to previous approaches, instead of solving the problem as change detection (using pre and post-event images), we model to identify destruction itself using a single post-event image. To overcome the challenge of collecting pixel-level ground truth data mostly used during training, we only assume image-level labels, representing either destruction is present (at any location) in a given image or not. The proposed attention-based mechanism learns to identify the image patches with destruction automatically under the sparsity constraint. Furthermore, to reduce false-positive and improve segmentation quality, a hard negative mining technique has been used. The implementation is mostly done in Keras.

Object Counting: Paper link,Paper link
Automatic counting of the number of the object of interest has many practical applications. In these projects, we attempt to address the challenging problem of counting built structures in satellite imagery. We propose a deep learning-based regression technique for counting built structures in satellite imagery. Our proposed framework intelligently combines features from different regions of satellite images using attention-based re-weighting techniques. Multiple parallel convolutional networks are designed to capture information at different granulates. These features are combined into the FusionNet which is trained to weigh features from different granularity differently, allowing us to predict a precise building count.

domain (developing regions). We initially align distribution maps across domains by aligning the output space distribution through adversarial loss. We then exploit counting consistency constraints, within-image count consistency, and across-image count consistency, to decrease the domain shift. Within-image consistency enforces that building count in the whole image should be greater than or equal to the count in any of its sub-image. Across-image consistency constraint enforces that if an image contains considerably more buildings than the other image, then their sub-images shall also have the same order. These two constraints encourage the behavior to be consistent across and within the images, regardless of the scale.

Few-Shot Images and Videos: Paper link, Paper link, Paper link
In several real-world applications, it is very difficult to obtain a sufficient number of training examples to train robust deep neural networks. To handle it, we proposed a few shot-learning approaches for human action classifications in videos. We explore two alternative data sources to improve aerial action classification when only a few training aerial examples are available. As a first data source, we resort to video games. We collect plenty of aerial game action videos using two gaming engines. For the second data source, we leverage conditional Wasserstein Generative Adversarial Networks (WGAN) to generate aerial features from ground videos. Given that both data sources have some limitations, e.g. game videos are biased towards specific actions categories (fighting, shooting, etc.,), and it is not easy to generate good discriminative GAN-generated features for all types of actions, we need to efficiently integrate two dataset sources with few available real aerial training videos. To address this challenge of the heterogeneous nature of the data, we propose to use a disjoint multitask learning framework. We feed the network with real and game, or real and GAN-generated data in an alternating fashion to obtain an improved action classifier.

Similarly, we employ single-shot learning for improving retinal image quality. Unlike typical deep learning-based approaches, our method does not require any training data. Instead, our Deep Image Prior (DIP)-based method can learn the underlying image prior while using a single degraded image. To perform retinal image enhancement, we frame it as a layer decomposition problem and investigate the use of two well-known analytical priors, i.e., dark channel prior (DCP) and bright channel prior (BCP) for atmospheric light estimation. We show that both the untrained neural networks and the pretrained neural networks can be used to generate an enhanced image while using only a single degraded image.

Invited Talks (Selected)

•'Real-world anomaly detection in Surveillance videos' at Lahore University of Management Sciences, Pakistan
•'Action Localization using Web Images' at Criminal Justice Department, University of Central Florida, USA
• 'Recognizing human actions on novel datasets' at Xerox Research Center, USA
• 'Weakly Labeled Action Detection' at Florida Institute of Technology, USA
• 'Automatic Action Annotations' at University of Florida, USA
• 'Weakly Labeled Human Action Recognition' at Punjab University College of Information Technology, Pakistan

Course Taught

Medical Image Computing (Ph.D. and MS), Advance Algorithm Analysis (Ph.D. and MS), Theory of Automata (BS)

Book Chapters

• **Waqas Sultani**, Qazi Ammar Ahmad, Chen Chen "Action Recognition in Real-World Videos", *Computer Vision: A Reference Guide, Springer Link, 2020*

• Sijie Zhu, Chen Chen, **Waqas Sultani** "Video Anomaly Detection for Smart Surveillance", *Computer Vision: A Reference Guide, Springer Link, 2020*

Academic Services

• Doctoral Program Review Committee memember
• Curriculum Design Program Review Committee member

Research Leadership

• Head of MedAI research group at Intelligent Machine lab, ITU, Punjab
• **PhD Students:** Current: 02 (Co-Supervision)
• **MS Students:** Graduated: 09, Current: 04

• **BS Students:** Graduated: 12, Current: 07

Community Services

•Reviwers of CVPR, ECCV, IEEE TIP, CVIU, IEEE PAMI
•Area Chair in ACM Multimedia Conference, 2020,2021, CVPR 2022, CVPR 2023
•Program Chair for Program Committee of WACV, HADCV workshop, 2021
•Area Chair for ACM Multimedia, 2020
•Program Chair for DL-HAU2020, ICPR, 2020
•Area Chair for ACM Multimedia 2020