POSTER SESSION BOOKLET

http://www.dmi.unict.it/icvss

University of Catania - University of Cambridge

International Computer Vision Summer School 2018

Computer Vision after Deep Learning

Sicily, 8 - 14 July 2018

International Computer Vision Summer School

Computer Vision is the science and technology of making machines that see. It is concerned with the theory, design and implementation of algorithms that can automatically process visual data to recognize objects, track and recover their shape and spatial layout.

The International Computer Vision Summer School - ICVSS was established in 2007 to provide both an objective and clear overview and an in-depth analysis of the state-of-the-art research in Computer Vision. The courses are delivered by world renowned experts in the field, from both academia and industry, and cover both theoretical and practical aspects of real Computer Vision problems.

The school is organized every year by University of Cambridge (Computer Vision and Robotics Group) and University of Catania (Image Processing Lab). The general entry point for past and future ICVSS editions is:

http://www.dmi.unict.it/icvss

ICVSS Poster Session

The International Computer Vision Summer School is especially aimed to provide a stimulating space for young researchers and Ph.D. Students. Participants have the possibility to present the results of their research, and to interact with their scientific peers, in a friendly and constructive environment.

This booklet contains the abstract of the posters accepted to ICVSS 2018.

Best Presentation Prize A subset of the submitted posters will be selected by the school committee for short oral presentation. A best presentation prize will be given to the best presentations selected by the school committee.

Scholarship A scholarship will be awarded to the best PhD student attending the school. The decision is made by the School Committee at the time of the School, taking into account candidates' CV, poster and oral presentation.

Sicily, June 2018

Roberto Cipolla
Sebastiano Battiato
Giovanni Maria Farinella

List of Posters [1]

1. WHICH BODY IS MINE? Sayed Mona-Ragab, Sim Terence, LIM Joo-Hwee, MA Keng-Teck
2. STEREO AND TOF DATA FUSION BY LEARNING FROM SYNTHETIC DATA Agresti G., Minto L., Marin G., Zanuttigh P.
3. AUTOMATED FACIAL SKIN ASSESSMENT Alarifi, J., Dancey, D., Yap, M.H.
4. DEEP SEGMENTATION NETWORK FOR BRAIN CONNECTIVITY DISSECTION Astolfi P., Sona D., Avesani P.
5. PERCEPTION AND ACTION IN PERIPERSONAL SPACE: A COMPARISON BETWEEN VIDEO AND OPTICAL SEE-THROUGH AUGMENTED REALITY DEVICES Ballestin G, Solari F, Chessa M
6. NATURAL INTERACTION IN AN IMMERSIVE VIRTUAL ENVIRONMENT Bassano C, Solari F, Chessa M
7. DEEP-TRACT: LEARNING TO TRACK BIOLOGICAL NEURAL PATHWAYS WITH ARTIFICIAL RECURRENT NEURAL NETS Benou I., Riklin Raviv T.
8. CODING KENDALL'S SHAPE TRAJECTORIES FOR 3D ACTION RECOGNITION Amor Ben Tanfous, Hassen Drira, Boulbaba Ben Amor
9. ONLINE PHOTOMETRIC CALIBRATION OF AUTO EXPOSURE VIDEO Bergmann P., Wang R., Cremers D.
10. UNSUPERVISED METRIC LEARNING FOR ACTIVITY ANALYSIS Brattoli B., Bchler U., Ommer B.
11. MEMORY BASED ONLINE LEARNING OF DEEP REPRESENTATIONS FROM VIDEO STREAMS Bruni M., Pernici F., Bartoli F., Del Bimbo A.

[1] Posters are ordered by surname of the speaker. Each poster is identified by a number. The page of a poster in this booklet corresponds with the ID of the poster.

12. LSTM SELF-SUPERVISION FOR DETAILED BEHAVIOR ANALYSIS Bchler U., Brattoli B., Ommer B.

13. RECOGNITION SELF-AWARENESS FOR ACTIVE OBJECT RECOGNITION Roberti A., Carletti M., Setti F. Cristani M.

14. ONE SHOT COGNITIVELY PLAUSIBLE LEARNING OF IMAGE-WORD ASSOCIATIONS Cerrato Mattia, Esposito Roberto, Gliozzi Valentina

15. SOUND CHARACTERIZATION VIA RATIONAL WAVELET TRANSFORM Bruni V., Della Cioppa L., Vitulano D.

16. THE TUM VI BENCHMARK FOR VISUAL-INERTIAL ODOMETRY Schubert D., Goll T., Demmel N., Usenko V., Stckler J., Cremers D.

17. SUPERPOINT: SELF-SUPERVISED INTEREST POINT DETECTION AND DESCRIPTION DeTone D., Malisiewicz T., Rabinovich A.

18. MACHINE LEARNING TECHNIQUES FOR EMOTIONS RECOGNITION Bertacchini N1., De Pietro N2.

19. LAYERED DEPTH PREDICTION FROM A SINGLE IMAGE Dhamo H., Tateno K., Laina I., Navab N., Tombari F.

20. SCENE ADAPTATION FOR SEMANTIC SEGMENTATION USING ADVERSARIAL LEARNING Di Mauro D, Furnari A, Patan G, Battiato S, Farinella GM

21. SPARSE FEATURE REGRESSION FOREST FOR REAL-TIME CAMERA RELOCALIZATION Duong N-D., Soladie C., Kacete A., Richard P-Y., Royan Jrme

22. LEARNING TO DETECT AND TRACK VISIBLE AND OCCLUDED BODY JOINTS IN A VIRTUAL WORLD Fabbri Matteo

23. FULL 3D LAYOUT RECONSTRUCTION FROM ONE SINGLE 360 IMAGE Fernandez Labrador Clara, Perez Yus Alejandro, Lopez Nicolas Gonzalo, Guerrero Jose J

24. VIDEO-REALISTIC EXPRESSIVE AUDIO-VISUAL SPEECH SYNTHESIS FOR THE GREEK LANGUAGE Filntisis P.P., Katsamanis A., Tsiakoulis P., Maragos P.

25. GENERAL SEMANTIC SEGMENTATION OF BONE AND CARDIAC TISSUE Fitzpatrick James, Trojan Lorenzo

26. SCALING EGOCENTRIC VISION: THE EPIC-KITCHENS DATASET Damen D., Doughty H., Farinella G. M., Fidler S., Furnari A., Kazakos E., Moltisanti D., Munro J., Perrett T., Prince W., Wray M.

27. ANOMALY DETECTION FOR SURVEILLANCE SYSTEM Ganokratanaa T., Aramvith S.

28. RESISTING ADVERSARIAL ATTACKS USING GAUSSIAN MIXTURE VARIATIONAL AUTOENCODERS Ghosh P., Losalka A., Black M.

29. UNDERSTANDING DEEP ARCHITECTURES BY VISUAL SUMMARIES Godi M., Carletti M., Aghaei M., Giuliari F., Cristani M.

30. LEARNING FROM #BARCELONA INSTAGRAM DATA: A LOCALS VS TOURIST ANALYSIS Gmez R., Karatzas D., Gmez L., Gibert J.

31. FULLY AUTOMATIC EXOSOMES SEGMENTATION IN TRANSMISSION ELECTRON MICROSCOPY IMAGES Gmez-de-Mariscal E., Maka M., Kotrbov A., Pospchalov V., Matula P., Muoz-Barrutia A.

32. SOE-NET: AN ANALYTICALLY DEFINED CONVNET WITH APPLICATION TO DYNAMIC TEXTURE RECOGNITION Hadji I., Wildes R.

33. INFINITE BRAIN MR IMAGES: PGGAN-BASED DATA AUGMENTATION FOR TUMOR DETECTION Han C., Hayashi H., Rundo L., Araki R., Nagano Y., Furukawa Y., Mauri G., Nakayama H.

34. SALIENCY PREDICTION, VISUALISATION AND APPLICATION Sen He, Nicolas Pugeault

35. GAMMALEARN: DEEP LEARNING APPLIED TO THE CHERENKOV TELESCOPE ARRAY (CTA) Jacquemont M.

36. MULTISCALE DESIGN RECOGNITION: CHALLENGES AND OPPORTUNITIES Jain A. and Kerne A.

37. CONDITIONAL IMAGE GENERATION FOR LEARNING THE STRUCTURE OF VISUAL OBJECTS Jakab T., Gupta A., Bilen H., Vedaldi A.

38. IMPLICIT REPRESENTATION FOR 3D SHAPE INFERENCE Pontes J. K., Michalkiewicz M., Sridharan S., Mahsa B., Fookes C., Eriksson A.

39. DEEP DRONE RACING: LEARNING AGILE FLIGHT IN DYNAMIC ENVIRONMENTS Kaufmann E., Loquercio A., Ranftl R., Dosovitskiy A., Koltun V., Scaramuzza D.

40. A DATA SET FOR PLACE RECOGNITION IN GARDENS Leyva Vallina M., Strisciuglio N., Petkov N.

41. IMPROVING SEMANTIC SEGMENTATION WITH A CARTOGRAPHIC PRIOR Loukkal A., Fremont V., Grandvalet Y., Li Y.

42. ADAPTABLE, SELF-SIMILARITY COUNTING NETWORKS Lu E., Xie W., Zisserman A.

43. DEEP MOTION SALIENCY DETECTION IN IMAGE SEQUENCES Maczyta L., Bouthemy P., Le Meur, O.

44. ITERATIVELY TRAINED INTERACTIVE SEGMENTATION Mahadevan S., Voigtlaender P., Leibe B.

45. SPORT ACTION RECOGNITION WITH SIAMESE SPATIO-TEMPORAL CNN, APPLICATION TO TABLE TENNIS Martin P.-E., Benois-Pineau J., Pteri R., Morlier J.

46. ON TRAINING TARGETS AND OBJECTIVE FUNCTIONS FOR DEEP LEARNING-BASED AUDIO-VISUAL SPEECH ENHANCEMENT Michelsanti D., Tan Z.-H., Sigurdsson S., Jensen J.

47. BONNET: AN OPEN-SOURCE TRAINING AND DEPLOYMENT FRAMEWORK FOR SEMANTIC SEGMENTATION IN ROBOTICS USING CNNS Milioto A., Stachniss C.

48. OPTIMAL STRUCTURED LIGHT A LA CARTE Mirdehghan P., Chen W., Kutulakos K.N.

49. PREDICTION OF ROAD ACTORS' BEHAVIOUR FOR AN AUTONOMOUS VEHICLE Mukherjee S., Wallace A., Wang S.

50. WEBLY SUPERVISED LEARNING FOR SKIN LESION CLASSIFICATION Navarro F., Conjeti S., Tombari F., Navab N.

51. GANS TO SYNTHETICALLY AUGMENT DATA FOR DEEP LEARNING BASED IMAGE SEGMENTATION Neff Thomas, Payer Christian, Stern Darko, Urschler Martin

52. FIRST STEPS TOWARDS A DEEP LEARNING HAND GESTURE RECOGNITION SET UP FOR COLLABORATIVE ROBOTS Nuzzi C., Pasinetti S., Lancini M., Docchio F., Sansoni G.

53. INDOOR LOCALIZATION FROM EGOCENTRIC IMAGES Orlando S, Furnari A, Battiato S, Farinella GM

54. ON THE PREDICTION OF SOCIAL IMAGE POPULARITY DYNAMICS Ortis A, Farinella GM, Battiato S

55. DEEP SKIN DETECTION ON LOW RESOLUTION GRAYSCALE IMAGES Paracchini Marco, Villa Federica, Marcon Marco

56. ROBUST SCENE GEOMETRY RECOVERY FROM EGO-CAMERAS FOR NAVIGATION Patra S., Banerjee S., Kalra P., Arora C.

57. A PREDICTIVE MODEL OF MEASUREMENT UNCERTAINTY FOR THE VELODYNE VLP-16 LIDAR Pntek Q., Allouis T., Strauss O., Fiorio C.

58. LEARNING DEEP FEATURES FROM ACOUSTIC IMAGES Perez AF., Morerio P., Murino V.

59. GENERATIVE ADVERSARIAL NETWORK FOR VIDEO FRAME INTERPOLATION Daniele Perlo, Andrea Forgione, Marco Grangetto

60. SEMANTIC HEAD SEGMENTATION IN CROWDED ENVIRONMENT Liciotti D., Paolanti M., Pietrini R., Frontoni E., Zingaretti P.

61. LEARNING 'OPEN' FROM 'CLOSE' Price W., Damen, D.

62. RECURRENT COLOR CONSTANCY Yanlin QianKe Chen, Jarno Nikkanen, Joni-Kristian Kmrinen, Jiri Matas

63. EGOCENTRIC VISITORS LOCALIZATION IN CULTURAL SITES Ragusa F, Furnari A, Battiato S, Signorello G, Farinella GM

64. MOVING AWAY FROM IMAGES - LEARNING ON POINT CLOUDS Rist C., Tzk A.

65. HUMAN ACTIVITY RECOGNITION IN AUTONOMOUS VEHICLES Roitberg A., Stiefelhagen R.

66. VISUAL SLAM BASED MARKERLESS INSIDE-OUT TRACKING Ruhkamp P.

67. SEGMENTATION-BASED WORD SPOTTING IN HANDWRITTEN DOCUMENTS Rusakov E., Rothacker L., Mo H., Fink G. A.

68. UNSUPERVISED LEARNING OF SENSORIMOTOR AFFORDANCES BY STOCHASTIC FUTURE PREDICTION Rybkin O.*, Pertsch K.*, Jaegle A., Derpanis K. and Daniilidis K.

69. UNSUPERVISED LEARNING OF SENSORIMOTOR AFFORDANCES BY STOCHASTIC FUTURE PREDICTION Rybkin O.*, Pertsch K.*, Jaegle A., Derpanis K. and Daniilidis K.

70. HOUSE SEGMENTATION WITH VARIATIONAL LEVEL SETS Sakeena M. , Zeppelzauer M.

71. POWER LAW SCALING OF TEST ERROR VERSUS NUMBER OF TRAINING IMAGES PER CLASS FOR A CONVOLUTIONAL NEURAL NETWORK IN MNIST HANDWRITTEN DIGITS AND MNIST FASHION ACCESSORIES DATASETS Sala V.

72. KEYPOINT LEARNING FOR 3D POSE ESTIMATION Saleh

73. EYE-TRACKED ANNOTATED DATA FOR SUPERVISED MACHINE LEARNING Samiei S, Rasti P, Ahmad A, Richard P, Belin E, Rousseau D.

74. ADAPTIVE QUANTILE SPARSE IMAGE PRIOR FOR INVERSE IMAGING PROBLEMS Schirrmacher Franziska, Khler Thomas, Riess Christian

75. MOBILE AUGMENTED REALITY FOR DENTISTRY Srs G., Lancelle, M., Degen N., Mrzinger R.

76. EGOCENTRIC SHOPPING CART LOCALIZATION Spera E, Furnari A, Battiato S, Farinella GM

77. INTERACTION RECOGNITION WITH DIVIDED ACTIVATIONS Stergiou A., Poppe R., Veltkamp R.

78. DIGITAL CAMERA IDENTIFICATION: SOME CRITICAL ISSUES IN EXISTING MODELS Bruni V., Tartaglione M., Vitulano D.

79. YOLO-REID: FAST SIMULTANEOUS PEOPLE DETECTION AND RE-IDENTIFICATION Van Ranst W., Goedem T.

80. PANSHARPENING BY CNN VITALE N1,FERRAIOLI N2,SCARPA N3

81. DIRECT SPARSE VISUAL-INERTIAL ODOMETRY USING DYNAMIC MARGINALIZATION von Stumberg L., Usenko V., Cremers D.

82. REAL-TIME FULLY INCREMENTAL SCENE UNDERSTANDING Wald J., Tateno K., Sturm J., Navab N., Tombari F.

83. RANGE SENSOR PERCEPTION FOR AUTOMATED DRIVING Wirges S.

84. GATED COMPLEX RECURRENT NEURAL NETWORKS Wolter Moritz, Yao Angela

85. ZERO-SHOT LEARNING Xian Yongqin

86. FUTURE PERSON LOCALIZATION IN FIRST-PERSON VIDEOS Yagi T., Mangalam K., Yonetani R., Sato Y.

87. UNSUPERVISED LEARNING OF OPTICAL FLOW VIA CLASSICAL OBJECTIVE FUNCTIONS Yu Jason, Derpanis Konstantinos

88. STEREO COMPUTATION FOR LOW-RESOLUTION THERMAL IMAGES Zoetgnande Yannick, Dillenseger Jean-Louis

89. PL-SLAM: STEREO SLAM FROM POINTS AND LINES Gomez-Ojeda R., Zuiga-Nol D., Moreno F.A., Scaramuzza D., Gonzalez-Jimenez J.

WHICH BODY IS MINE?

Sayed Mona-Ragab, Sim Terence, LIM Joo-Hwee, MA Keng-Teck

Abstract: We propose a new research problem: head-body matching. Given an image of a person's head, we match it with his body (headless) image. We propose a dual-pathway framework which computes head and body features independently and learns the correlation between such features. We use anthropometric features and state-of-the-art deep-CNN features in our framework. We demonstrate the usefulness of our framework with two novel applications: head/body recognition, and T-shirt sizing from a head image.

Contact: a0135563@u.nus.edu

Presentation Type: Poster

Date: Monday 9 July 2018

Time: 21:30

Poster Session: 1

STEREO AND TOF DATA FUSION BY LEARNING FROM SYNTHETIC DATA

Agresti G., Minto L., Marin G., Zanuttigh P.

Abstract: ToF and stereo vision systems are range imaging devices with complementary strengths and issues. We present a framework for data fusion enforcing the local consistency of the depth data, by considering the confidence maps of the sensors that are jointly estimated using a CNN. The training is carried out on a synthetic dataset. We show how this approach can generalize to real world data and increase the depth estimation accuracy on both synthetic and real data with state-of-the-art performances.

Contact: gianluca.agresti@phd.unipd.it

Presentation Type: Poster

Date: Monday 9 July 2018

Time: 21:30

Poster Session: 1

AUTOMATED FACIAL SKIN ASSESSMENT

Alarifi, J., Dancey, D., Yap, M.H.

Abstract: In computer vision, some applications have been developed to achieve an objective measure of skin properties such as wrinkles and spots. Recent developments of deep Convolutional Neural Networks have increased the efficiency of age estimation, but it is not clear what is the role of skin properties in such method. This work aims to investigate the automated computerised methods for facial skin assessment. First, we demonstrate the ability of various automated computerised methods in skin types classification. Second, we illustrate the face features for age estimation from human and machine perspectives. We conclude by providing future direction of the work.

Contact: Jhan.s.alarifi@mmu.ac.uk

Presentation Type: Poster

Date: Monday 9 July 2018

Time: 21:30

Poster Session: 1

DEEP SEGMENTATION NETWORK FOR BRAIN CONNECTIVITY DISSECTION

Astolfi P., Sona D., Avesani P.

Abstract: Recent studies proved the presence of ROIs in the brain white matter, called stems, able to fully characterise certain bundles of brain connections, called tracts. In this work we focus on the IFOF tract, and we provide a method to automatically dissect it by segmenting its corresponding stem ROI. We re-implement and adapt a powerful deep network, RefineNet, to segment the stem ROI of new subjects. Finally, we combine the predicted stem ROIs with specific cortical regions to dissect IFOF tracts.

Contact: pietroastolfi92@gmail.com

Presentation Type: Poster

Date: Monday 9 July 2018

Time: 21:30

Poster Session: 1

PERCEPTION AND ACTION IN PERIPERSONAL SPACE: A COMPARISON BETWEEN VIDEO AND OPTICAL SEE-THROUGH AUGMENTED REALITY DEVICES

Ballestin G, Solari F, Chessa M

Abstract: This work analyzes how we perceive the peripersonal space when involved in a reaching task in an Augmented Reality (AR) environment. We consider the difference between two different AR wearable devices: optical see-through (OST) and video see-through (VST) head-mounted displays (HMDs). Our results revealed a better depth estimation, thus a more precise interaction, when using the OST device, which also revealed a lower impact on eye strain and fatigue.

Contact: giorgio.ballestin@gmail.com

Presentation Type: Poster

Date: Monday 9 July 2018

Time: 21:30

Poster Session: 1

NATURAL INTERACTION IN AN IMMERSIVE VIRTUAL ENVIRONMENT

Bassano C, Solari F, Chessa M

Abstract: This work investigates how people interact in non-immersive and immersive virtual reality environments (VE), by accomplishing selection and manipulation tasks, when different interaction devices (touchful and touchless) and simple feedbacks (visual and audio) are introduced. Results show that only in non-immersive 2D VE, where no 3D stereoscopic cues are provided, feedbacks positively affect performances. Moreover, both touchful and touchless modalities have been effective and appreciated. This analysis may play a role in the design of natural and ecological interfaces.

Contact: chiara.bassano@dibris.unige.it

Presentation Type: Poster

Date: Monday 9 July 2018

Time: 21:30

Poster Session: 1

DEEP-TRACT: LEARNING TO TRACK BIOLOGICAL NEURAL PATHWAYS WITH ARTIFICIAL RECURRENT NEURAL NETS

Benou I., Riklin Raviv T.

Abstract: Tractography is a 3D modeling technique for visualizing neural tracts, used in surgical planning and brain studies. Goal: Develop a deep-learning framework for probabilistic tractography from raw diffusion MRI (dMRI) data. Key Concept: Recurrent neural net is trained to estimate orientation distribution functions (ODFs) of the brain neural pathways. Tractography is then performed by sequentially sampling from these distributions at test time. Experiments are based on the ISMRM challenge data.

Contact: benoui@post.bgu.ac.il

Presentation Type: Poster

Date: Monday 9 July 2018

Time: 21:30

Poster Session: 1

CODING KENDALL'S SHAPE TRAJECTORIES FOR 3D ACTION RECOGNITION

Amor Ben Tanfous, Hassen Drira, Boulbaba Ben Amor

Abstract: Suitable shape representations as well as their temporal evolution, termed trajectories, often lie to non-linear manifolds. This puts an additional constraint (i.e., non-linearity) in using conventional machine learning techniques. To handle this problem in the context of action recognition, we propose a novel human actions representation based on an intrinsic sparse coding of skeletal shapes on the Kendall's shape space. This allows to map skeletal trajectories from a non-linear space to discriminative sparse time-series in Euclidean space.

Contact: omar.bentanfous@imt-lille-douai.fr

Presentation Type: Poster

Date: Monday 9 July 2018

Time: 21:30

Poster Session: 1

ONLINE PHOTOMETRIC CALIBRATION OF AUTO EXPOSURE VIDEO

Bergmann P., Wang R., Cremers D.

Abstract: We present an approach for full photometric calibration of arbitrary auto exposure video sequences, recovering exposure times of consecutive frames, the camera response function and the attenuation factors of the sensor irradiance due to vignetting in real-time. The calibration can significantly enhance the performance of direct methods in computer vision that work only on pixel intensities. Physical access to the camera itself is not required.

Contact: paulbergmann21@gmail.com

Presentation Type: Poster

Date: Monday 9 July 2018

Time: 21:30

Poster Session: 1

UNSUPERVISED METRIC LEARNING FOR ACTIVITY ANALYSIS

Brattoli B., Bchler U., Ommer B.

Abstract: Understanding human activity and being able to explain it in detail surpasses mere action classification by far in both complexity and value. The challenge is thus to describe an activity on the basis of its most fundamental constituents, the individual postures and their distinctive transitions. We propose three completely unsupervised deep learning procedures without requiring pre-trained networks, predefined body models or keypoints

Contact: biagio.brattoli@gmail.com

Presentation Type: Poster

Date: Monday 9 July 2018

Time: 21:30

Poster Session: 1

MEMORY BASED ONLINE LEARNING OF DEEP REPRESENTATIONS FROM VIDEO STREAMS

Bruni M., Pernici F., Bartoli F., Del Bimbo A.

Abstract: We present a novel online unsupervised method for face identity learning from video streams. The method exploits deep face descriptors together with a memory based learning mechanism that takes advantage of the temporal coherence of visual data. We introduce a discriminative descriptor matching solution based on Reverse Nearest Neighbour and a forgetting strategy that detect redundant descriptors and discard them appropriately while time progresses.

Contact: matteo.bruni@unifi.it

Presentation Type: Poster

Date: Monday 9 July 2018

Time: 21:30

Poster Session: 1

LSTM SELF-SUPERVISION FOR DETAILED BEHAVIOR ANALYSIS

Bchler U., Brattoli B., Ommer B.

Abstract: We utilize self-supervision to automatically learn accurate posture and behavior representations for analyzing motor function. Our model involves the following crucial elements: (i) OOI detection based on an FCN is initialized solely using motion information, (ii) a novel self-supervised training of LSTMs using only temporal permutation yields a detailed representation of behavior, and (iii) back- propagation of this representation also improves the description of individual postures.

Contact: uta.buechler@iwr.uni-heidelberg.de

Presentation Type: Poster

Date: Monday 9 July 2018

Time: 21:30

Poster Session: 1

RECOGNITION SELF-AWARENESS FOR ACTIVE OBJECT RECOGNITION

Roberti A., Carletti M., Setti F. Cristani M.

Abstract: We propose an active object recognition framework that introduces the recognition self-awareness, aka RA-POMDP, which is an intermediate level of reasoning to decide which views to cover during object exploration: * train a multi-view deep 3D object classifier on depth images; * a 3D dense saliency volume is generated by fusing together single-view visualization maps; * the volume is injected in the observation model of a POMDP [1].

Contact: marco.carletti@univr.it

Presentation Type: Poster

Date: Monday 9 July 2018

Time: 21:30

Poster Session: 1

ONE SHOT COGNITIVELY PLAUSIBLE LEARNING OF IMAGE-WORD ASSOCIATIONS

Cerrato Mattia, Esposito Roberto, Gliozzi Valentina

Abstract: At around 18 months of age,children demonstrate a dramatic increase in the rate of word acquisition.To explain how this kind of efficiency is possible,cognitivists theorize that 1)children need few object presentations;2)they tend to assign the same word to similar objects;3)they understand whether words refer to whole objects or their parts.Our model extends previous work by Mayor and Plunkett [1] and uses Deep Networks,SOMs and Hebbian Learning to perform word learning under these constraints.

Contact: cerrato@di.unito.it

Presentation Type: Poster

Date: Monday 9 July 2018

Time: 21:30

Poster Session: 1

SOUND CHARACTERIZATION VIA RATIONAL WAVELET TRANSFORM

Bruni V., Della Cioppa L., Vitulano D.

Abstract: In the field of voice processing, the Mel-cepstrum transform combines two elements: the logarithmically spaced Mel scale, which models the human auditory system; and the cepstrum transform, which allows separation between excitation and resonances. The aim of our research is to mimic the Mel scale using discrete rational wavelet transform (RADWT), more flexible in terms of frequency spacing than the discrete wavelet transform (DWT), to obtain the Mel frequency cepstral coefficients (MFCC).

Contact: lorenzo.dellacioppa@uniroma1.it

Presentation Type: Poster

Date: Monday 9 July 2018

Time: 21:30

Poster Session: 1

THE TUM VI BENCHMARK FOR VISUAL-INERTIAL ODOMETRY

Schubert D., Goll T., Demmel N., Usenko V., Stckler J., Cremers D.

Abstract: We propose the TUM VI benchmark, a novel dataset with 28 diverse sequences in different scenes for evaluating VI odometry. It includes camera images with 1024x1024 resolution at 20Hz, high dynamic range and photometric calibration, as well as IMU measurements at 200Hz, all time-synchronized in hardware. Moreover, we provide accurate pose ground truth from a motion capture system at 120Hz at the start and end of each sequence and evaluate two state-of-the-art VI odometry approaches.

Contact: demmeln@in.tum.de

Presentation Type: Poster

Date: Monday 9 July 2018

Time: 21:30

Poster Session: 1

SUPERPOINT: SELF-SUPERVISED INTEREST POINT DETECTION AND DESCRIPTION

DeTone D., Malisiewicz T., Rabinovich A.

Abstract: This paper presents a self-supervised framework for training interest point detectors and descriptors suitable for a large number of multiple-view geometry problems in computer vision. Our model jointly computes pixel-level keypoints and descriptors in one forward pass. We introduce Homographic Adaptation, an approach for boosting interest point detection repeatability and performing cross-domain adaptation. Our model gives rise to state-of-the-art homography estimation results on HPatches.

Contact: ddetone@magicleap.com

Presentation Type: Poster

Date: Monday 9 July 2018

Time: 21:30

Poster Session: 1

MACHINE LEARNING TECHNIQUES FOR EMOTIONS RECOGNITION

Bertacchini N1., De Pietro N2.

Abstract: Photos from Cohn-Kanade database of subjects expressing 6 emotions (Anger, Disgust, Fear, Happiness, Sadness and Surprise) are selected. These photos are analyzed using two machine learning techniques in order to investigate the best method to recognize emotions automatically. The methods used are Microsoft Azure and Wolfram Mathematica. The results obtained showed that the first method is more accurate, because it identified a lot of emotions compared to the second method.

Contact: michela.depietro@unical.it

Presentation Type: Poster

Date: Monday 9 July 2018

Time: 21:30

Poster Session: 1

LAYERED DEPTH PREDICTION FROM A SINGLE IMAGE

Dhamo H., Tateno K., Laina I., Navab N., Tombari F.

Abstract: With this work, we extend the conventional depth map prediction to the regression of a specific data representation called Layered Depth Image (LDI), which contains information about the occluded regions in the reference frame and can fill in occlusion gaps in case of small view changes. We propose a novel approach based on Convolutional Neural Networks (CNNs) to jointly predict depth maps and foreground separation masks used to condition Generative Adversarial Networks (GANs) for hallucinating plausible color and depths in the initially occluded areas. We demonstrate the effectiveness of our approach for novel scene view synthesis from a single image.

Contact: helisa.dhamo@tum.de

Presentation Type: Poster

Date: Monday 9 July 2018

Time: 21:30

Poster Session: 1

SCENE ADAPTATION FOR SEMANTIC SEGMENTATION USING ADVERSARIAL LEARNING

Di Mauro D, Furnari A, Patan G, Battiato S, Farinella GM

Abstract: We have proposed a method to perform scene adaptation using adversarial learning. The method improves the generalization of a semantic segmentation network by enforcing the reconstruction of the input images from the generated semantic segmentation masks. Experiments show that the proposed method greatly reduces over-fitting in both point of view and scene adaptation. Future work will explicitly consider the geometric alignment of source and target scenes to improve the results in the case of point of view adaptation.

Contact: dimauro@dmi.unict.it

Presentation Type: Poster

Date: Monday 9 July 2018

Time: 21:30

Poster Session: 1

SPARSE FEATURE REGRESSION FOREST FOR REAL-TIME CAMERA RELOCALIZATION

Duong N-D., Soladie C., Kacete A., Richard P-Y., Royan Jrme

Abstract: Camera relocalization is needed in several applications such as augmented reality or robot navigation. However, it is still challenging to have both a real-time and accurate method. We present our hybrid method combining machine learning approach and geometric approach for real-time camera relocalization from a single RGB image. We propose a sparse feature regression forest to improve the machine learning part. The results indicate that our method is the only real-time hybrid method (50ms per frame) while being as accurate as the best state-of the-art methods hybrid methods.

Contact: Nam-Duong.DUONG@b-com.com

Presentation Type: Poster

Date: Monday 9 July 2018

Time: 21:30

Poster Session: 1

LEARNING TO DETECT AND TRACK VISIBLE AND OCCLUDED BODY JOINTS IN A VIRTUAL WORLD

Fabbri Matteo

Abstract: Multi-People Tracking in an open-world setting requires a special effort in precise detection. For the purpose, we propose a deep network architecture that jointly extracts people body parts and associates them across short temporal spans. Our model explicitly deals with occluded body parts, by hallucinating plausible solutions of not visible joints. We also created the vastest CG dataset for people tracking in urban scenarios by exploiting a photorealistic videogame of human body parts for people tracking in urban scenarios. Our architecture trained on virtual data exhibits good generalization capabilities also on public real tracking benchmarks.

Contact: matteo.fabbri@unimore.it

Presentation Type: Poster

Date: Monday 9 July 2018

Time: 21:30

Poster Session: 1

FULL 3D LAYOUT RECONSTRUCTION FROM ONE SINGLE 360 IMAGE

Fernandez Labrador Clara, Perez Yus Alejandro, Lopez Nicolas Gonzalo, Guerrero Jose J

Abstract: We propose a novel entire pipeline which converts 360o panoramas into flexible, closed, 3D reconstructions of the rooms represented in the images. Key ideas: 1. Explotation of deep learning techniques combined with geometric reasoning to obtain structural lines. 2. New Normal Map for the hypotheses evaluation step. 3. Final closed, 3D room reconstructions faithful to the actual shapes.

Contact: cfernandez@unizar.es

Presentation Type: Poster

Date: Monday 9 July 2018

Time: 21:30

Poster Session: 1

VIDEO-REALISTIC EXPRESSIVE AUDIO-VISUAL SPEECH SYNTHESIS FOR THE GREEK LANGUAGE

Filntisis P.P., Katsamanis A., Tsiakoulis P., Maragos P.

Abstract: Expressive talking agents are vital for natural human-computer interaction. In this paper we propose two deep neural network (DNN) architectures for generating expressive talking heads and show that they outperform traditional hidden Markov models (HMM) and Unit Selection (US) methods. We also employ HMM adaptation and interpolation and show that they can effectively be used for adapting a talking head to a target facial expression and generating faces with mixed facial expressions.

Contact: filby@central.ntua.gr

Presentation Type: Poster

Date: Monday 9 July 2018

Time: 21:30

Poster Session: 1

GENERAL SEMANTIC SEGMENTATION OF BONE AND CARDIAC TISSUE

Fitzpatrick James, Trojan Lorenzo

Abstract: In this work we aim to develop an automatic segmentation algorithm that can quickly and accurately classify anatomical features of interest from medical images. Weshowthatgeneral pixelwise classication of cortical and trabecular bone from CT scans can be achieved with high accuracy, even when training on a small number of manually-segmented images. We will now focus on development of an algorithm to segment cardiac soft tissue from MR images.

Contact: j.fitzpatrick@axial3d.com

Presentation Type: Poster

Date: Monday 9 July 2018

Time: 21:30

Poster Session: 1

SCALING EGOCENTRIC VISION: THE EPIC-KITCHENS DATASET

Damen D., Doughty H., Farinella G. M., Fidler S., Furnari A., Kazakos E., Moltisanti D., Munro J., Perrett T., Prince W., Wray M.

Abstract: First-person vision is gaining interest as it offers a unique viewpoint on people's interaction with objects, their attention, and even intention. However, progress in this challenging domain has been relatively slow due to the lack of sufficiently large datasets.In this paper, we introduce EPIC-KITCHENS, a large-scale egocentric video benchmark recorded by 32 participants in their native kitchen environments. Our videos depict nonscripted daily activities: we simply asked each participant to start recording every time they entered their kitchen. Recording took place in 4 cities (in North America and Europe) by participants belonging to 10 different nationalities, resulting in highly diverse kitchen habits and cooking styles. Our dataset features 55 hours of video consisting of 11.5M frames, which we densely labeled for a total of 39.6K action segments and 454.2K object bounding boxes. Our annotation is unique in that we had the participants narrate their own videos (after recording), thus reflecting true intention, and we crowd-sourced ground-truths based on these. We describe our object, action and anticipation challenges, and evaluate several baselines over two test splits, seen and unseen kitchens.

Contact: furnari@dmi.unict.it

Presentation Type: Poster

Date: Monday 9 July 2018

Time: 21:30

Poster Session: 1

ANOMALY DETECTION FOR SURVEILLANCE SYSTEM

Ganokratanaa T., Aramvith S.

Abstract: Anomaly detection is useful for real-time surveillance system. Current works mostly run offline and struggle with object localization. Thus, we propose the real time anomaly detection using GANs with optical flow. The experiment shows the fast and accurate object localization results.

Contact: charismasbunny@gmail.com

Presentation Type: Poster

Date: Monday 9 July 2018

Time: 21:30

Poster Session: 1

RESISTING ADVERSARIAL ATTACKS USING GAUSSIAN MIXTURE VARIATIONAL AUTOENCODERS

Ghosh P., Losalka A., Black M.

Abstract: DNN classifiers are vulnerable to fooling and adversarial samples". These thus far have been considered separately by all defense mechanisms. Our method[0] addresses them both. We modify a VAE to have a Gaussian mixture prior for the latent variable. This enables us to perform selective classification, leading to rejection of adversarial samples, with an optional reclassification step using our generative model. Our method can also be trained in a semi-supervised setting.

Contact: partha.ghosh@tuebingen.mpg.de

Presentation Type: Poster

Date: Monday 9 July 2018

Time: 21:30

Poster Session: 1

UNDERSTANDING DEEP ARCHITECTURES BY VISUAL SUMMARIES

Godi M., Carletti M., Aghaei M., Giuliari F., Cristani M.

Abstract: Visualization techniques investigate the recurrent visual patterns exploited by deep networks for object classification. Unfortunately, no effort has been spent in showing that these techniques are effective in leading researchers to univocal and exhaustive explanations. Our approach leads to a group of summaries, each one by regions focusing on a particular semantic part that the network has exploited with a given class. The approach[1] is effective is transmitting this information to the user, in a fast, effective and coherent way, and it is shown very robust to misinterpretations.

Contact: marco.godi@univr.it

Presentation Type: Poster

Date: Monday 9 July 2018

Time: 21:30

Poster Session: 1

LEARNING FROM #BARCELONA INSTAGRAM DATA: A LOCALS VS TOURIST ANALYSIS

Gmez R., Karatzas D., Gmez L., Gibert J.

Abstract: We gather Instagram data (image-captions pairs) related to barcelona and perform a language separate analysis. We learn relations between words, images and neighbourhoods in a self-supervised way. We show that the language separate treatment can be extrapolated to a tourist vs locals analysis, and that tourism is reflected in Social Media at a neighbourhood level.

Contact: raul.gomez@cvc.uab.es

Presentation Type: Poster

Date: Monday 9 July 2018

Time: 21:30

Poster Session: 1

FULLY AUTOMATIC EXOSOMES SEGMENTATION IN TRANSMISSION ELECTRON MICROSCOPY IMAGES

Gmez-de-Mariscal E., Maka M., Kotrbov A., Pospchalov V., Matula P., Muoz-Barrutia A.

Abstract: Exosomes are nano-scale cell-derived extracellular vesicles, involved in the intercellular communication. Exosomes quantification is currently done manually by biologists and its automation will help them to remarkably progress in their research. We present the Fully Residual Unet for the segmentation of exosomes in Transmission Electron Microscope images and the Radon transform properties to separate clusters. An accuracy over 80% and 2s processing time for 2048x2048 pixels image are achieved.

Contact: esgomezm@pa.uc3m.es

Presentation Type: Poster

Date: Monday 9 July 2018

Time: 21:30

Poster Session: 1

SOE-NET: AN ANALYTICALLY DEFINED CONVNET WITH APPLICATION TO DYNAMIC TEXTURE RECOGNITION

Hadji I., Wildes R.

Abstract: SOE-Net combines the benefits of the multilayer architecture of ConvNets and a more controlled approach to spacetime image analysis. Unlike most ConvNets no learning is involved; instead, all components are defined analytically. This approach makes it possible to understand what information is being extracted at each stage and minimizes heuristic choices in design.

Contact: hadjisma@cse.yorku.ca

Presentation Type: Poster

Date: Monday 9 July 2018

Time: 21:30

Poster Session: 1

INFINITE BRAIN MR IMAGES: PGGAN-BASED DATA AUGMENTATION FOR TUMOR DETECTION

Han C., Hayashi H., Rundo L., Araki R., Nagano Y., Furukawa Y., Mauri G., Nakayama H.

Abstract: Due to the lack of available annotated medical images, accurate computer-assisted diagnosis requires intensive Data Augmentation (DA) techniques, such as geometry/intensity transformations of original images; however, those transformed images intrinsically have a similar distribution to the original ones, leading to limited performance improvement. To fill the data lack in the real image distribution, we synthesize brain contrast-enhanced Magnetic Resonance (MR) images—realistic but completely different from the original ones—using Generative Adversarial Networks (GANs). This study exploits Progressive Growing of GANs (PGGANs), a multi-stage generative training method, to generate original-sized 256 256 MR images for convolutional neural network-based brain tumor detection, which is challenging via conventional GANs; difficulties arise due to unstable GAN training with high resolution and a variety of tumors in size, location, shape, and contrast. Our preliminary results show that this novel PGGAN-based DA method can achieve promising performance improvement, when combined with classical DA, in tumor detection and also in other medical imaging tasks.

Contact: han@nlab.ci.i.u-tokyo.ac.jp

Presentation Type: Poster

Date: Monday 9 July 2018

Time: 21:30

Poster Session: 1

SALIENCY PREDICTION, VISUALISATION AND APPLICATION

Sen He, Nicolas Pugeault

Abstract: An essential mechanism that allows the human visual system to process information in real time is its capacity to selectively focus attention on parts of the scene. This process has been extensively studied by Psychologists to discover which visual patterns capture human attention. In this work, we shed light on predicting where we will look given a static image and analysing what kind of stimuli attract our visual attention. More specifically, we developed a simple yet powerful deep model to predict the the salient regions in the image that attract our visual attention, and by visualising the developed model, we analysed what patterns ours eyes are attuned to. Furthermore, We apply this attention mechanism to image style transfer, and show the promising results by this application.

Contact: sh752@exeter.ac.uk

Presentation Type: Poster

Date: Monday 9 July 2018

Time: 21:30

Poster Session: 1

GAMMALEARN: DEEP LEARNING APPLIED TO THE CHERENKOV TELESCOPE ARRAY (CTA)

Jacquemont M.

Abstract: The Cherenkov Telescope Array (CTA) is the next generation ground-based observatory for -ray astronomy. It will be used to study -ray sources, allowing to better understand the Universe. It will generate petabytes of data per year, leading to big data challenges. The GammaLearn project proposes to apply Deep Learning as a part of the analysis of this huge amount of data. Its goal is to sep- arate the photons from cosmic particles, and reconstruct the photons parameters, from noisy unconventional images (hexago- nal grid, non rectangular shape).

Contact: mikael.jacquemont@lapp.in2p3.fr

Presentation Type: Poster

Date: Monday 9 July 2018

Time: 21:30

Poster Session: 1

MULTISCALE DESIGN RECOGNITION: CHALLENGES AND OPPORTUNITIES

Jain A. and Kerne A.

Abstract: Recognition algorithms based on computer vision fall short when applied to multiscale designs e.g. infographics created using Photoshop or Illustrator, presentations created with Prezi, and exhibits created with LiveMch. The primary reason is that the algorithms are not specifically designed to operate on content organized at multiple zoom levels. To deal with novel challenges posed by the problem domain, we propose using multimodal analysis and combining design product and process semantics.

Contact: ajit@ecologylab.net

Presentation Type: Poster

Date: Monday 9 July 2018

Time: 21:30

Poster Session: 1

CONDITIONAL IMAGE GENERATION FOR LEARNING THE STRUCTURE OF VISUAL OBJECTS

Jakab T., Gupta A., Bilen H., Vedaldi A.

Abstract: We consider the problem of unsupervised learning of landmarks. An image of an object is conditionally generated from another one by providing the generator with a keypoint-like representation extracted from the target image through a bottleneck. The representation is encouraged to distil the object geometry, different between source and target, while the appearance, shared between both, is read off from the source alone. Our approach outperforms state-of-the-art methods on a variety of datasets.

Contact: tomj@robots.ox.ac.uk

Presentation Type: Poster

Date: Monday 9 July 2018

Time: 21:30

Poster Session: 1

IMPLICIT REPRESENTATION FOR 3D SHAPE INFERENCE

Pontes J. K., Michalkiewicz M., Sridharan S., Mahsa B., Fookes C., Eriksson A.

Abstract: How to effectively represent 3D shape in a learning framework still remains an open research problem. Recent works have been relying on volumetric, point cloud and mesh representations, but these approaches typically suffer from issues of computational complexity, unordered data, and lack of finer geometry. Here, we instead propose the use of implicit functions to represent shape for 3D shape inference. We show that this formulation will address many of the shortcomings of existing methods in an elegant way, while still compactly and accurately encoding highly detailed geometry.

Contact: jhonykaesemodel@gmail.com

Presentation Type: Poster

Date: Monday 9 July 2018

Time: 21:30

Poster Session: 1

DEEP DRONE RACING: LEARNING AGILE FLIGHT IN DYNAMIC ENVIRONMENTS

Kaufmann E., Loquercio A., Ranftl R., Dosovitskiy A., Koltun V., Scaramuzza D.

Abstract: We propose a learning-based approach to autonomous drone racing. Our approach combines a convolutional neural network with state-of-the-art trajectory generation. The CNN produces a direction and speed from a single image, which are then used to generate a short trajectory segment. Our method does not require any explicit map of the environment and runs fully onboard. We extensively test and compare the precision and robustness of the approach in simulation and in the real world.

Contact: ekaufmann@ifi.uzh.ch

Presentation Type: Poster

Date: Monday 9 July 2018

Time: 21:30

Poster Session: 1

A DATA SET FOR PLACE RECOGNITION IN GARDENS

Leyva Vallina M., Strisciuglio N., Petkov N.

Abstract: We consider place recognition as the task of identifying whether a pair of images under different viewpoints depict the same place, based on visual cues only. Recently, the use of robots in agriculture and automatic gardening created new challenges due to the highly textured garden-like environments that robots have to interact with [1]. In this work we propose a new data set for place recognition in gardens and benchmark several state of the art holistic descriptors.

Contact: m.leyva.vallina@rug.nl

Presentation Type: Poster

Date: Monday 9 July 2018

Time: 21:30

Poster Session: 1

IMPROVING SEMANTIC SEGMENTATION WITH A CARTOGRAPHIC PRIOR

Loukkal A., Fremont V., Grandvalet Y., Li Y.

Abstract: We describe 3 approaches to inject a cartographic prior in semantic segmentation CNNs :

- in the fitting criteria, as a pair-wise potentials in CRFs, - as an additional output of the CNN, - as an additional input of a CNN.

The best approach improves the IoU, compared to state-of-the-art CNN, for the pedestrians, vehicles and traffic signs classes by respectively 4%, 1.58% and 8.93% on the CARLA simulator.

Contact: abdelhak.loukkal@gmail.com

Presentation Type: Poster

Date: Monday 9 July 2018

Time: 21:30

Poster Session: 1

ADAPTABLE, SELF-SIMILARITY COUNTING NETWORKS

Lu E., Xie W., Zisserman A.

Abstract: Nearly all existing counting methods are designed for a specific object class. We propose a class-agnostic counting model which exploits the property of image self-similarity, an important prior not yet leveraged in counting work. Treating counting as a matching problem, we pretrain a base model on object tracking data, and adapt it to novel counting domains with very few examples and trainable parameters. We demonstrate our method on different benchmarks and achieve competitive results.

Contact: erika@robots.ox.ac.uk

Presentation Type: Poster

Date: Monday 9 July 2018

Time: 21:30

Poster Session: 1

DEEP MOTION SALIENCY DETECTION IN IMAGE SEQUENCES

Maczyta L., Bouthemy P., Le Meur, O.

Abstract: The aim of this work is to study motion saliency in image sequences. Here, salient elements are defined as elements whose image motion departs from their spatio-temporal context, with a camera that can also be mobile.

We have defined so far deep learning architectures to temporally detect motion saliency, i.e., to classify every image of the video sequence as dynamically salient or not, meaning that it contains dynamic salient objects or not.

Contact: leo.maczyta@inria.fr

Presentation Type: Poster

Date: Monday 9 July 2018

Time: 21:30

Poster Session: 1

ITERATIVELY TRAINED INTERACTIVE SEGMENTATION

Mahadevan S., Voigtlaender P., Leibe B.

Abstract: For the task of object segmentation, manually labelling data is very expensive, and hence interactive methods are needed. Following recent approaches, we develop an interactive object segmentation system which uses user input, in the form of clicks, as the input to a convolutional network. We propose a new iterative training strategy, where clicks are added based on the errors in the current prediction, and show that it improves the results over the state-of-the-art.

Contact: sabarinath.mahadevan@rwth-aachen.de

Presentation Type: Poster

Date: Monday 9 July 2018

Time: 21:30

Poster Session: 1

SPORT ACTION RECOGNITION WITH SIAMESE SPATIO-TEMPORAL CNN, APPLICATION TO TABLE TENNIS

Martin P.-E., Benois-Pineau J., Pteri R., Morlier J.

Abstract: Our goal is recognition of actions in sports with the aim of improving athletes' performances. We apply our method to strokes recognition in table tennis. Twenty stroke classes and a rejection class are considered accordingly to the table tennis rules. The objective is to help the teachers to focus on particular strokes performed by students. Our model performed 91.4% of accuracy against 43.1% for the state of the art method.

Contact: pierre.etienne.martin.sendin@gmail.com

Presentation Type: Poster

Date: Tuesday 10 July 2018

Time: 21:30

Poster Session: 2

ON TRAINING TARGETS AND OBJECTIVE FUNCTIONS FOR DEEP LEARNING-BASED AUDIO-VISUAL SPEECH ENHANCEMENT

Michelsanti D., Tan Z.-H., Sigurdsson S., Jensen J.

Abstract: Audio-visual speech enhancement is the task of improving speech quality and intelligibility in a noisy environment given the video of a talker. In this study, we explore the impact that video frames from the talker's lips have for different training targets and objective functions of a deep learning framework.

Contact: danmi@es.aau.dk

Presentation Type: Poster

Date: Tuesday 10 July 2018

Time: 21:30

Poster Session: 2

BONNET: AN OPEN-SOURCE TRAINING AND DEPLOYMENT FRAMEWORK FOR SEMANTIC SEGMENTATION IN ROBOTICS USING CNNS

Milioto A., Stachniss C.

Abstract: Semantic scene understanding is important for several robotic applications. Convolutional Neural Networks (CNNs) have become a standard for this task. Available frameworks have a steep learning curve for a non-expert in machine learning. Our framework allows to quickly implement new models and datasets to approach new problems. Bonnet has a stable C++ deployment library and ROS-compatible interface for different hardware accelerators.

Contact: amilioto@uni-bonn.de

Presentation Type: Poster

Date: Tuesday 10 July 2018

Time: 21:30

Poster Session: 2

OPTIMAL STRUCTURED LIGHT A LA CARTE

Mirdehghan P., Chen W., Kutulakos K.N.

Abstract: We propose a framework for automatically generating structured light patterns for active stereo triangulation of a static scene. Unlike existing approaches that use predetermined patterns and reconstruction algorithms tied to them, we generate patterns on the fly in response to generic specifications. Our pattern sequences are specifically optimized to minimize the expected rate of correspondence errors under those specifications for an unknown scene, and are coupled to a sequence-independent algorithm for per-pixel disparity estimation.

Contact: parsa@cs.toronto.edu

Presentation Type: Poster

Date: Tuesday 10 July 2018

Time: 21:30

Poster Session: 2

PREDICTION OF ROAD ACTORS' BEHAVIOUR FOR AN AUTONOMOUS VEHICLE

Mukherjee S., Wallace A., Wang S.

Abstract: Despite recent progress in vehicle autonomy, significant improvement is needed in decision making, predicting the behaviour of other road users within the field of interest. Human behaviour is the outcome of decisions made by the driver. In our autonomous approach, high level classification (e.g. left, right turn, lane change etc.) are made currently by a decision tree and low level predictions by KNN based trajectory clustering.

Contact: saptarshim11@gmail.com

Presentation Type: Poster

Date: Tuesday 10 July 2018

Time: 21:30

Poster Session: 2

WEBLY SUPERVISED LEARNING FOR SKIN LESION CLASSIFICATION

Navarro F., Conjeti S., Tombari F., Navab N.

Abstract: Within medical imaging, manual curation of sufficient well-labeled samples is cost, time and scale-prohibitive. To improve the representativeness of the training dataset, for the

rst time, we present an approach to utilize large amounts of freely available web data through web-crawling. To handle noise and weak nature of web annotations, we propose a two-step transfer learning based training process with a robust loss function, termed as Webly Supervised Learning (WSL) to train deep models for the task. We also leverage search by image to improve the search speci

fity of our web-crawling and reduce cross domain noise. Within WSL, we explicitly model the noise structure between classes and incorporate it to selectively distill knowledge from the web data during model training. To demonstrate improved performance due to WSL, we benchmarked on a publicly available 10-class

ne-grained skin lesion classi

cation dataset and report a signi

ficant improvement of top-1 classi

cation accuracy from 71.25 % to 80.53 % due to the incorporation of web-supervision.

Contact: fernando.navarro@tum.de

Presentation Type: Poster

Date: Tuesday 10 July 2018

Time: 21:30

Poster Session: 2

GANS TO SYNTHETICALLY AUGMENT DATA FOR DEEP LEARNING BASED IMAGE SEGMENTATION

Neff Thomas, Payer Christian, Stern Darko, Urschler Martin

Abstract: Especially for small datasets, where deep learning methods often struggle to train effectively, data augmentation is commonly used to improve the training process. We evaluate a data augmentation strategy utilizing Generative Adversarial Networks (GANs) and perform an extensive evaluation on two segmentation tasks. By learning to generate additional, labeled data, our goal is to remove the need for prior domain knowledge to find suitable data augmentation parameters, leading to a more efficient training process.

Contact: thomas.neff@student.tugraz.at

Presentation Type: Poster

Date: Tuesday 10 July 2018

Time: 21:30

Poster Session: 2

FIRST STEPS TOWARDS A DEEP LEARNING HAND GESTURE RECOGNITION SET UP FOR COLLABORATIVE ROBOTS

Nuzzi C., Pasinetti S., Lancini M., Docchio F., Sansoni G.

Abstract: We present a smart hand gesture recognition experimental set up for collaborative robots, using a Faster R-CNN object detector to find the accurate position of the hands in the RGB images. We used MATLAB to code the detector and a purposely designed function for the prediction phase, necessary for detecting static gestures in the way we have defined them. We performed a number of experiments with four different datasets to evaluate the performances of the system in different situations. We also tested the system in real-time, achieving good performances that can lead to real-time human-robot interaction, being the inference time around 0.2 seconds.

Contact: c.nuzzi@unibs.it

Presentation Type: Poster

Date: Tuesday 10 July 2018

Time: 21:30

Poster Session: 2

INDOOR LOCALIZATION FROM EGOCENTRIC IMAGES

Orlando S, Furnari A, Battiato S, Farinella GM

Abstract: Image-Based Localization (IBL) is a challenging problem. The aim of IBL is to understand where a picture has been acquired by obtaining the position and orientation of the camera with respect to a 3D world. IBL is useful to overcome classic positioning systems, such as GPS, which are not suitable in indoor environments. In this paper we investigate an Image-Based localization approach based on a deep learning architecture. We propose an automatic way to generate a 6DOF labeled image dataset of indoor environments by using the Unity3d engine in order to train an IBL model. The proposed IBL approach achieves promising results in preliminary experiments.

Contact: santi.orlando@unict.it

Presentation Type: Poster

Date: Tuesday 10 July 2018

Time: 21:30

Poster Session: 2

ON THE PREDICTION OF SOCIAL IMAGE POPULARITY DYNAMICS

Ortis A, Farinella GM, Battiato S

Abstract: We consider the challenge of forecasting the engagement score reached by social images over time. We call this task "Popularity Dynamic Prediction" . The goal is to predict the engagement score dynamic over a period of time (e.g., 30 days) by exploiting visual and social features. To this aim, we propose a benchmark dataset that consists of 20K Flickr images labelled with their engagement scores (i.e., views, comments and favorites) in a period of 30 days from the upload in the social platform. For each image, the dataset also includes user's and photo's social features that have been proven to have an influence on the image popularity on Flickr. The proposed dataset is publicly available for research purpose. We also present a method to address the aforementioned problem. Our approach is able to forecast the daily number of views reached by a photo posted on Flickr for a period of 30 days, by exploiting features extracted from the post.

Contact: alessandro.ortis@gmail.com

Presentation Type: Poster

Date: Tuesday 10 July 2018

Time: 21:30

Poster Session: 2

DEEP SKIN DETECTION ON LOW RESOLUTION GRAYSCALE IMAGES

Paracchini Marco, Villa Federica, Marcon Marco

Abstract: In this work we present a facial skin detection method, based on a deep learning architecture, that is able to precisely label each pixel of a given image depicting a face as skin or not. In particular the proposed method is able to work on low resolution grayscale images (even 64x64 pixel) depicting non aligned faces in presence of general lighting illumination, facial expressions, object occlusions and regardless of the gender, age and ethnicity of the subject.

Contact: marcobrando.paracchini@polimi.it

Presentation Type: Poster

Date: Tuesday 10 July 2018

Time: 21:30

Poster Session: 2

ROBUST SCENE GEOMETRY RECOVERY FROM EGO-CAMERAS FOR NAVIGATION

Patra S., Banerjee S., Kalra P., Arora C.

Abstract: A variety of applications demand navigation in an unknown environment using ego-cameras. For reliable analysis, we need to estimate the location of the robot/car, the environment around, and a navigable space infront. Depending upon the applications, the granularity and accuracy of each of this estimations become very important. In this work, we estimate these parameters robustly and efficiently in presence of wild motions and low parallax and further use it for navigation on road. We prove the accuracy of each of these parameters on different benchmark datasets.

Contact: suvam@cse.iitd.ac.in

Presentation Type: Poster

Date: Tuesday 10 July 2018

Time: 21:30

Poster Session: 2

A PREDICTIVE MODEL OF MEASUREMENT UNCERTAINTY FOR THE VELODYNE VLP-16 LIDAR

Pntek Q., Allouis T., Strauss O., Fiorio C.

Abstract: A key feature for multi-sensor fusion is the ability to associate, to each measurement, an estimate of its uncertainty. We aim at developing a point-to-pixel association based on UAV-borne LiDAR and camera data to build colored digital elevation models. We propose an uncertainty prediction model dedicated to LiDAR systems that we validate using a reference-free evaluation method based on two criteria: consistency and specificity. Experiment on the VLP-16 LiDAR shows the model returns coherently pessimistic predictions with a resolving power upper bounded by 2 cm at a distance of 5 m.

Contact: quentin.pentek@lirmm.fr

Presentation Type: Poster

Date: Tuesday 10 July 2018

Time: 21:30

Poster Session: 2

LEARNING DEEP FEATURES FROM ACOUSTIC IMAGES

Perez AF., Morerio P., Murino V.

Abstract: In this project it is explored the utilization of deep learning models to learn rich feature representations from acoustic data collected with the hybrid sensor proposed in [1], also known as DualCam. We aim at leveraging the pixel-level association and the natural synchronization between video and audio data provided by the sensor to address tasks such as acoustic scene classification and acoustic object localization under a multi-modal and self-supervised learning setting.

Contact: andres.perez@iit.it

Presentation Type: Poster

Date: Tuesday 10 July 2018

Time: 21:30

Poster Session: 2

GENERATIVE ADVERSARIAL NETWORK FOR VIDEO FRAME INTERPOLATION

Daniele Perlo, Andrea Forgione, Marco Grangetto

Abstract: Video streaming is the most widely diffuse type of entertainment nowadays. Today the amount of data transmitted for video streams is growing faster than ever. So, due to the increase the performance as trade off between transmission bitrate and video quality, several techniques have been proposed to enhance the frame rate by image prediction, beside image compression algorithms. We are developing a generative adversarial network in order to predict a video frame from its neighbors, in according to the objects translation and per-pixel gradient.

Contact: dperlo@unito.it

Presentation Type: Poster

Date: Tuesday 10 July 2018

Time: 21:30

Poster Session: 2

SEMANTIC HEAD SEGMENTATION IN CROWDED ENVIRONMENT

Liciotti D., Paolanti M., Pietrini R., Frontoni E., Zingaretti P.

Abstract: Detecting and tracking people is a challenging task in a persistent crowded environment, defined as at least 3 people per square meter are passing in the area of the camera, with some cases of physical contacts be- tween people entering the area of in- terest (i.e.retail, airport, station, etc.) for human behaviour analysis of secu- rity purposes. This work introduces an approach to detect people in cases of heavy occlusions based on CNNs for semantic segmentation using top- view visual data. The purpose is the design of a novel U-Net architec- ture, U-Net 3. The approach was ap- plied and tested on a new and public available dataset, TVHeads Dataset, consisting of depth images of people recorded from an RGB-D camera in- stalled in top-view configuration. Our variant outperforms baseline architec- tures while remaining computation- ally efficient at inference time. Results show high accuracy, demonstrating the effectiveness and suitability of our approach.

Contact: r.pietrini@pm.univpm.it

Presentation Type: Poster

Date: Tuesday 10 July 2018

Time: 21:30

Poster Session: 2

LEARNING 'OPEN' FROM 'CLOSE'

Price W., Damen, D.

Abstract: We investigate the role time plays in action recognition by conducting a study of action recognition under the reversal of time for the purposes of data augmentation and zero shot learning on the EPIC Kitchens dataset. We evaluate the idea using Temporal Segment Network using the batch-normalized Inception architecture.

Contact: will.price@bristol.ac.uk

Presentation Type: Poster

Date: Tuesday 10 July 2018

Time: 21:30

Poster Session: 2

RECURRENT COLOR CONSTANCY

Yanlin QianKe Chen, Jarno Nikkanen, Joni-Kristian Kmrinen, Jiri Matas

Abstract: We introduce a novel formulation of temporal color constancy which considers multiple frames preceding the frame for which illumination is estimated. We propose an end-to-end trainable recurrent color constancy network - the RCC-Net - which exploits convolutional LSTMs and a simulated sequence to learn compositional representations in space and time. We use a standard single frame color constancy benchmark, the SFU Gray Ball Dataset, which can be adapted to a temporal setting. Extensive experiments show that the proposed method consistently outperforms single-frame state-of-the-art methods and their temporal variants.

Contact: yanlin.qian@tut.fi

Presentation Type: Poster

Date: Tuesday 10 July 2018

Time: 21:30

Poster Session: 2

EGOCENTRIC VISITORS LOCALIZATION IN CULTURAL SITES

Ragusa F, Furnari A, Battiato S, Signorello G, Farinella GM

Abstract: We consider the problem of localizing visitors in a cultural site from egocentric (first person) images. Localization information can be useful both to assist the user during his visit (e.g., by suggesting where to go and what to see next) and to provide behavioral information to the manager of the cultural site (e.g., how much time has been spent by visitors at a given location? What has been liked most?). To tackle the problem, we collected a large dataset of egocentric videos using two cameras: a head-mounted HoloLens device and a chest-mounted GoPro. Each frame has been labeled according to the location of the visitor and to what he was looking at. The dataset is freely available in order to encourage research in this domain. The dataset is complemented with baseline experiments performed considering a state-of-the-art method for location-based temporal segmentation of egocentric videos. Experiments show that compelling results can be achieved to extract useful information for both the visitor and the site-manager.

Contact: francesco.ragusa@unict.it

Presentation Type: Poster

Date: Tuesday 10 July 2018

Time: 21:30

Poster Session: 2

MOVING AWAY FROM IMAGES - LEARNING ON POINT CLOUDS

Rist C., Tzk A.

Abstract: DNN architectures and methods for object detection, segmentation, optical flow have primarily been invented and advanced on image data. To achieve the topmost performance in the context of environment perception with LiDAR for intelligent vehicles these methods need to get transfered to the domain of point clouds. Non-uniform sampling density, a varying number of points in each sample and permutation invariance impose challenges for the DNN architecture not found in the image domain.

Contact: christoph_bernd.rist@daimler.com

Presentation Type: Poster

Date: Tuesday 10 July 2018

Time: 21:30

Poster Session: 2

HUMAN ACTIVITY RECOGNITION IN AUTONOMOUS VEHICLES

Roitberg A., Stiefelhagen R.

Abstract: Understanding the situation inside intelligent cars is crucial for intuitive and safe human-vehicle cooperation. We aim to bridge the gap between the theoretical approaches for activity recognition and practical applications for driver monitoring with real-life constraints, such as low illumination, scarce training data and dynamic open-set environment. We identify key design requirements for such in-cabin recognition systems and present an overview of our on-going work on fine-grained activity detection under realistic driving conditions.

Contact: alina.roitberg@kit.edu

Presentation Type: Poster

Date: Tuesday 10 July 2018

Time: 21:30

Poster Session: 2

VISUAL SLAM BASED MARKERLESS INSIDE-OUT TRACKING

Ruhkamp P.

Abstract: Outside-in tracking, here in clinical scenarios, is limited by e.g. the line-of-sight problem. We evaluate the accuracy and applicability of markerless visual-SLAM based inside-out tracking. A small stereo camera is mounted on top of an ultrasound probe and a trans-rectal ultrasound (TRUS) examination of a prostate phantom is performed for qualitative 3D tomographic reconstruction. Accuracy is evaluated by a quantitative analysis between different methods against robotic ground-truth.

Contact: p.ruhkamp@framos.com

Presentation Type: Poster

Date: Tuesday 10 July 2018

Time: 21:30

Poster Session: 2

SEGMENTATION-BASED WORD SPOTTING IN HANDWRITTEN DOCUMENTS

Rusakov E., Rothacker L., Mo H., Fink G. A.

Abstract: Given a user defined query, the task of word spotting is to retrieve a list containing word images that are relevant with respect to the query. Typically word spotting methods rank all retrieved word images from a given document collection by a certain criterion and sort them by their similarities. These queries can either be word images, defined by a user cropping a snipped from a document page or defining a word string which needs to be retrieved. As dictionaries can contain many thousands of words, present-day word recognition methods require attribute-based classification approaches. Hence, Almazan et al. [2] proposed a word string embedding called Pyramidal Histogram of Characters (PHOC), representing the character occurences as binary attributes. Retrieval is performed by comparing the predicted PHOC vectors using a certain distance measure, for example the cosine similarity. In this work [1], the cosine similarity is replaced by a probabilistic comparison of similarities. Lampert et al. [5] proposed a method for a similarity measure called Direct Attribute Prediction (DAP), comparing the posteriors instead of distances.

Contact: eugen.rusakov@tu-dortmund.de

Presentation Type: Poster

Date: Tuesday 10 July 2018

Time: 21:30

Poster Session: 2

UNSUPERVISED LEARNING OF SENSORIMOTOR AFFORDANCES BY STOCHASTIC FUTURE PREDICTION

Rybkin O.*, Pertsch K.*, Jaegle A., Derpanis K. and Daniilidis K.

Abstract: Intelligent perception must capture not only a scenes static content, but also its affordances: how an agent's actions can affect the scene. We propose an unsupervised method to learn an environment's sensorimotor affordances. We use a recurrent latent variable that is

(i) minimal in sensitivity to static content and (ii) compositional in nature.

We show these two properties are sufficient to induce representations that are reusable across different scenes with shared degrees of freedom.

Contact: oleh@seas.upenn.edu

Presentation Type: Poster

Date: Tuesday 10 July 2018

Time: 21:30

Poster Session: 2

UNSUPERVISED LEARNING OF SENSORIMOTOR AFFORDANCES BY STOCHASTIC FUTURE PREDICTION

Rybkin O.*, Pertsch K.*, Jaegle A., Derpanis K. and Daniilidis K.

Abstract: Intelligent perception must capture not only a scenes static content, but also its affordances: how an agent's actions can affect the scene. We propose an unsupervised method to learn an environment's sensorimotor affordances. We use a recurrent latent variable that is

(i) minimal in sensitivity to static content and (ii) compositional in nature.

We show these two properties are sufficient to induce representations that are reusable across different scenes with shared degrees of freedom.

Contact: oleh@seas.upenn.edu

Presentation Type: Poster

Date: Tuesday 10 July 2018

Time: 21:30

Poster Session: 2

HOUSE SEGMENTATION WITH VARIATIONAL LEVEL SETS

Sakeena M. , Zeppelzauer M.

Abstract: Deep learning (DL) approaches proved to be effective in many computer vision applications. We investigate the capabilities of DL in the field of real estate image analysis which is an emerging and challenging research topic. We propose the fusion of DL and variational level sets for segmentation of houses from real estate (RE) images. We evaluate our approach on a dataset provided by RE experts. Our promising results show the significance of DL in image analysis for RE applications.

Contact: muntahasakeena@hotmail.com

Presentation Type: Poster

Date: Tuesday 10 July 2018

Time: 21:30

Poster Session: 2

POWER LAW SCALING OF TEST ERROR VERSUS NUMBER OF TRAINING IMAGES PER CLASS FOR A CONVOLUTIONAL NEURAL NETWORK IN MNIST HANDWRITTEN DIGITS AND MNIST FASHION ACCESSORIES DATASETS

Sala V.

Abstract: Accuracy and size of the training set are key parameters for the industrial application of image classification algorithms based on convolutional neural networks. In this work, the same deep neural network was trained on subsets of variable size of both the MNIST and MNIST fashion datasets. The accuracy appears to be well described by the sum of two components: the first scales as the inverse of the square root of the number of the training images per class while the second appears constant.

Contact: vittorio.sala@imagesspa.it

Presentation Type: Poster

Date: Tuesday 10 July 2018

Time: 21:30

Poster Session: 2

KEYPOINT LEARNING FOR 3D POSE ESTIMATION

Saleh

Abstract: In computer vision there are tasks such as object detection or tracking which are beholden to keypoint detection. Here we propose a pipeline to learn keypoints from rendered CAD models, which can be used in tracking and AR. A network is used to generate keypoint locations as heat maps. We implement a Siamese network to input stereo pairs and use the learned keypoint estimations for 3D pose estimation using marker point cloud registration. Finally, we transfer the model to directly estimate pose.

Contact: m.saleh@tum.de

Presentation Type: Poster

Date: Tuesday 10 July 2018

Time: 21:30

Poster Session: 2

EYE-TRACKED ANNOTATED DATA FOR SUPERVISED MACHINE LEARNING

Samiei S, Rasti P, Ahmad A, Richard P, Belin E, Rousseau D.

Abstract: An open problem in robotized agriculture is to detect weed in dense culture. This problem can be addressed with computer vision and machine learning. The bottleneck of supervised approach lay in the manual annotation of training images. We propose to speed up this process with eye-tracked annotated images which are at least 30 times faster. Such annotations also come with less resolution and accuracy. We investigate the impact on the detection of the weeds.

Contact: salma.samiei@etud.univ-angers.fr

Presentation Type: Poster

Date: Tuesday 10 July 2018

Time: 21:30

Poster Session: 2

ADAPTIVE QUANTILE SPARSE IMAGE PRIOR FOR INVERSE IMAGING PROBLEMS

Schirrmacher Franziska, Khler Thomas, Riess Christian

Abstract: Inverse problems can solve computer vision related tasks. An appropriate prior to model natural images is a crucial step when designing well-performing methods. We propose the Adaptive Quantile Sparse Image (AQuaSI) prior based on the weighted quantile filter. The prior can be easily plugged into existing methods and its efficacy is demonstrated by means of joint upsampling.

Contact: franziska.schirrmacher@fau.de

Presentation Type: Poster

Date: Tuesday 10 July 2018

Time: 21:30

Poster Session: 2

MOBILE AUGMENTED REALITY FOR DENTISTRY

Srs G., Lancelle, M., Degen N., Mrzinger R.

Abstract: Digital technologies have the potential to disrupt many traditional industries including dentistry. Software becomes the main communication technology and a useful tool for showing patients the possibilities for enhancing their smiles. Despite the need for processing the visual content, todays solutions do not correspond to the state of the art in computer vision. We demonstrate a mobile application that enables users to virtually try on new teeth and dental restorations in real time. Based on our proprietary visual computing technology stack built on machine learning and dento-facial analytics, we create realistic previews for patients, dental professionals, and the dental industry.

Contact: gabor.soros@kapanu.com

Presentation Type: Poster

Date: Tuesday 10 July 2018

Time: 21:30

Poster Session: 2

EGOCENTRIC SHOPPING CART LOCALIZATION

Spera E, Furnari A, Battiato S, Farinella GM

Abstract: This work investigates the new problem of image-based egocentric shopping cart localization in retail stores. The contribution of our work is two-fold. First, we propose a novel large-scale dataset for image-based egocentric shopping cart localization. The dataset has been collected using cameras placed on shopping carts in a large retail store. It contains a total of 19,531 image frames, each labelled with its six Degrees Of Freedom pose. We study the localization problem by analysing how cart locations should be represented and estimated, and how to assess the localization results. Second, we benchmark different image-based techniques to address the task. Specifically, we investigate two families of algorithms: classic methods based on image retrieval and emerging methods based on regression. Experimental results show that methods based on image retrieval largely outperform regression-based approaches. We also point out that deep metric learning techniques allow to learn better visual representations w.r.t. other architectures, and are useful to improve the localization results of both retrieval-based and regression-based approaches. Our findings suggest that deep metric learning techniques can help bridge the gap between retrieval-based and regression-based methods.

Contact: emilianospera@libero.it

Presentation Type: Poster

Date: Tuesday 10 July 2018

Time: 21:30

Poster Session: 2

INTERACTION RECOGNITION WITH DIVIDED ACTIVATIONS

Stergiou A., Poppe R., Veltkamp R.

Abstract: Interaction recognition considers the automated analysis of video sequences. The main challenges in this task are the variations in terms of durations, backgrounds, lighting, viewpoints and the overall people movements in videos. This project introduces the use of an Energy Based Model (EBM) as a method for allowing 3DCNNs to learn a robust representation of the entire sequence.

Contact: a.g.stergiou@uu.nl

Presentation Type: Poster

Date: Tuesday 10 July 2018

Time: 21:30

Poster Session: 2

DIGITAL CAMERA IDENTIFICATION: SOME CRITICAL ISSUES IN EXISTING MODELS

Bruni V., Tartaglione M., Vitulano D.

Abstract: Source camera identification from digital images is a challenging task in digital forensics. Existing methods are based on the extraction of a specific noise component, the photo-response non-uniformity (PRNU), which is a stochastic signature of the device. This poster focuses on two critical issues which generally occur in the existing models: (i) a lot of information is required in order to extract a reliable PRNU; (ii) camera identification from post-processed images is more critical.

Contact: michela.tartaglione@sbai.uniroma1.it

Presentation Type: Poster

Date: Tuesday 10 July 2018

Time: 21:30

Poster Session: 2

YOLO-REID: FAST SIMULTANEOUS PEOPLE DETECTION AND RE-IDENTIFICATION

Van Ranst W., Goedem T.

Abstract: In this work we combine detection and re-id into one single pass network based on the YOLO architecture. We propose a network that can do re-id simultaneously with detection and classification. The effect of our modification has only a negligible impact on detection accuracy, and adds the calculation of re-id vectors at virtually no cost. The resulting re-id vector is strong enough to be used in speed sensitive applications which can benefit from an additional re-id vector.

Contact: wiebevr@gmail.com

Presentation Type: Poster

Date: Tuesday 10 July 2018

Time: 21:30

Poster Session: 2

PANSHARPENING BY CNN

VITALE N1,FERRAIOLI N2,SCARPA N3

Abstract: This presentation summarizes the research activity, carried out in the first year of my PhD program, which is focused on pansharpening of multiresolution satellite images such as those provided by WorldView, GeoEye or Ikonos missions. Due to technological constraints, such satellites have been conceived to provide composite images made of a single panchromatic (PAN) band at a very high spatial resolution (VHR) and a lower resolution multispectral (MS) image, trading-off between spatial and spectral resolutions. The goal of pansharpening is the fusion of the MS and PAN components aimed to get a spatial-spectral full-resolution datacube by mixing the spatial details conveyed by PAN with the spectral information provided by MS. Recently, inspired by the successful results obtained on super-resolution tasks by using CNNs, Masi et. al. [1] proposed the first CNN solution to pansharpening achieving state-of-the-art-results. Following this research line, I have proposed [2, 3] several modifications to the baseline [1], the most relevant being the use of residual learning, a different loss function, the introduction of a target-adaptive scheme. The experimental results, both numerically and visually, show a significant improvement over the state-of-the-art methods, including the baseline [1].

Contact: sergio.vitale@uniparthenope.it

Presentation Type: Poster

Date: Tuesday 10 July 2018

Time: 21:30

Poster Session: 2

DIRECT SPARSE VISUAL-INERTIAL ODOMETRY USING DYNAMIC MARGINALIZATION

von Stumberg L., Usenko V., Cremers D.

Abstract: We present VI-DSO. We propose a novel initialization strategy, where scale and gravity direction are jointly optimized after the initialization. In order to keep the system consistent we propose "dynamic marginalization. The idea is to adaptively employ marginalization strategies even in cases where certain variables undergo drastic changes. Evaluation shows that our method outperforms the state of the art.

Contact: stumberg@in.tum.de

Presentation Type: Poster

Date: Tuesday 10 July 2018

Time: 21:30

Poster Session: 2

REAL-TIME FULLY INCREMENTAL SCENE UNDERSTANDING

Wald J., Tateno K., Sturm J., Navab N., Tombari F.

Abstract: We propose an online RGB-D based scene understanding method for indoor scenes running in real-time on mobile devices. First, we incrementally reconstruct and segment a scene [1] by fusing segments obtained from the depth. To accomplish efficient semantic segmentation, we encode the segments with a fast incremental 3D descriptor and use a random forest to determine its semantic label. Our accuracy is comparable to state-of-the-art 3D scene understanding methods while being much more efficient.

Contact: johanna.wald@tum.de

Presentation Type: Poster

Date: Tuesday 10 July 2018

Time: 21:30

Poster Session: 2

RANGE SENSOR PERCEPTION FOR AUTOMATED DRIVING

Wirges S.

Abstract: Environment perception is a crucial component for the safe operation of automated vehicles. However, to deal with the amount of perceived information we require further segmentation, categorization and abstraction strategies.

In the following we represent the environment by top-view grid maps. Occupancy grid maps encode surface point positions and free-space from a point of view in a two-dimensional grid. As all traffic participants move on a common ground surface one might not require full 3D information but instead represent the scene on a 2D plane with obstacles occupying areas along the drivable path. The use of multi-layer grids has several advantages. They are well-suited for free-space estimation, sensor fusion and their organized 2D structure enables the use of efficient convolutional operations for deep learning in contrast to sparse point sets.

Here, we present a selection of methods intended to be a basis for further discussion. Please do not hesitate to contact me in-person or via email.

Contact: sascha.wirges@kit.edu

Presentation Type: Poster

Date: Tuesday 10 July 2018

Time: 21:30

Poster Session: 2

GATED COMPLEX RECURRENT NEURAL NETWORKS

Wolter Moritz, Yao Angela

Abstract: Complex numbers have long been favored for digital signal processing, yet complex representations rarely appear in deep learning architectures. RNNs, widely used to process time series and sequence information, could greatly benefit from complex representations. We present a novel complex gated recurrent cell. When used together with norm-preserving state transition matrices, our complex gated RNN exhibits excellent stability and convergence properties. We demonstrate competitive performance of our complex gated RNN on the synthetic memory and adding task, as well as on the real-world task of human motion prediction.

Contact: wolter@cs.uni-bonn.de

Presentation Type: Poster

Date: Tuesday 10 July 2018

Time: 21:30

Poster Session: 2

ZERO-SHOT LEARNING

Xian Yongqin

Abstract: Suffering from the extreme training data imbalance between seen and unseen classes, most of existing state-of-the-art approaches fail to achieve satisfactory results for the challenging generalized zero-shot learning task. To circumvent the need for labeled examples of unseen classes, we propose a novel generative adversarial network (GAN) that synthesizes CNN features conditioned on class-level semantic information, offering a shortcut directly from a semantic descriptor of a class to a class-conditional feature distribution. Our proposed approach, pairing a Wasserstein GAN with a classification loss, is able to generate sufficiently discriminative CNN features to train softmax classifiers or any multimodal embedding method. Our experimental results demonstrate a significant boost in accuracy over the state of the art on five challenging datasets – CUB, FLO, SUN, AWA and ImageNet – in both the zero-shot learning and generalized zero-shot learning settings.

Contact: yxian@mpi-inf.mpg.de

Presentation Type: Poster

Date: Tuesday 10 July 2018

Time: 21:30

Poster Session: 2

FUTURE PERSON LOCALIZATION IN FIRST-PERSON VIDEOS

Yagi T., Mangalam K., Yonetani R., Sato Y.

Abstract: We present a new task that predicts future locations of people observed in first-person videos (FPVs). To achieve this task, we make two key observations: a) FPVs typically involve significant ego-motion which greatly affects the location of the target person in future frames; b) FPVs often capture people up-close, making it easier to leverage target poses (e.g., where they look) for predicting their future locations. We incorporate these observations into a multi-stream CNN.

Contact: tyagi@iis.u-tokyo.ac.jp

Presentation Type: Poster

Date: Tuesday 10 July 2018

Time: 21:30

Poster Session: 2

UNSUPERVISED LEARNING OF OPTICAL FLOW VIA CLASSICAL OBJECTIVE FUNCTIONS

Yu Jason, Derpanis Konstantinos

Abstract: Optical flow is a core computer vision problem concerned with recovering the apparent motion of visual patterns in a set of temporal images, such as the consecutive frames of a video. Successful approaches using machine learning in the form of convnets to estimating optical flow rely on large annotated datasets that are expensive to curate. We propose an unsupervised approach based on classical objective functions to train a convnet end-to-end without the need for groundtruth optical flow.

Contact: jason.jiasheng.yu@ryerson.ca

Presentation Type: Poster

Date: Tuesday 10 July 2018

Time: 21:30

Poster Session: 2

STEREO COMPUTATION FOR LOW-RESOLUTION THERMAL IMAGES

Zoetgnande Yannick, Dillenseger Jean-Louis

Abstract: Stereo computation can be used for 3D vision. We have performed a stereo computation method for very low-resolution thermal images (80x60). A stereo calibration method is first proposed for such images. Once calibration is performed we extracted features using Phase congruency, then these features are matched. Due to low resolution, we performed a sub-pixel matching method using phase correlation in the Phase congruency magnitude space.

Contact: yannick.zoetgnande@etudiant.univ-rennes1.fr

Presentation Type: Poster

Date: Tuesday 10 July 2018

Time: 21:30

Poster Session: 2

PL-SLAM: STEREO SLAM FROM POINTS AND LINES

Gomez-Ojeda R., Zuiga-Nol D., Moreno F.A., Scaramuzza D., Gonzalez-Jimenez J.

Abstract: Traditional approaches to visual SLAM rely on point features to estimate the camera trajectory and build a map of the environment. In low-textured environments, it is often difficult to find reliable point features and hence the performance of such algorithms degrades. PL-SLAM is a stereo visual SLAM system that combines both points and line segments at all the instances of the process (visual odometry, keyframe selection, bundle adjustment, etc) to work robustly in a wider variety of scenarios. We also contribute with a loop closure procedure through a novel bag-of-words approach that exploits the combined descriptive power of the two kinds of features. An open source version of the PL-SLAM C++ code has been released for the benefit of the community: github.com/rubengooj/pl-slam

Contact: dzuniga@uma.es

Presentation Type: Poster

Date: Tuesday 10 July 2018

Time: 21:30

Poster Session: 2