

# ON THE IMPACT OF VARYING REGION PROPOSAL STRATEGIES FOR RAINDROP DETECTION AND CLASSIFICATION USING CONVOLUTIONAL NEURAL NETWORKS

Tiancheng Guo<sup>1</sup>, Samet Akcay<sup>1</sup>, Philip A. Adey<sup>1</sup>, Toby P. Breckon<sup>1,2</sup>

Department of {Computer Science<sup>1</sup> | Engineering<sup>2</sup>}, Durham University, UK.

## ABSTRACT

The presence of raindrop induced image distortion has a significant negative impact on the performance of a wide range of all-weather visual sensing applications including within the increasingly important contexts of visual surveillance and vehicle autonomy. A key part of this problem is robust raindrop detection such that the potential for performance degradation in effected image regions can be identified. Here we address the problem of raindrop detection in colour video imagery by considering three varying region proposal approaches with secondary classification via a number of novel convolutional neural network architecture variants. This is verified over an extensive dataset with in-frame raindrop annotation to achieve maximal 0.95 detection accuracy with minimal false positives compared to prior work. Our approach is evaluated under a range of environmental conditions typical of all-weather automotive visual sensing applications.

**Index Terms**— rain detection, raindrop distortion, all-weather computer vision, automotive vision, CNN

## 1. INTRODUCTION

Despite their potential impact across a wide range of computer vision applications, prior work in the detection and removal of raindrops in video imagery is limited [6, 22, 26]. Furthermore, as the range of automotive sensing applications such as sign detection [5] and feature-driven odometry [16, 17], in addition to generalised scene understanding [13, 2] and stereo-based sensing [11, 10], are becoming more prevalent in the automotive industry, all-weather operation is becoming a key topic for future vehicle autonomy. In addition, the impact of adverse weather conditions notably affects other common visual surveillance tasks [3, 4]. However, despite the potential impact of raindrop distortion upon such approaches, work on raindrop removal remains in its infancy [22, 27, 26] characterised by performance characteristics below what would be expected within contemporary scene understanding approaches (i.e. false negative : 15% - [27, 26], 7% - [26]).

In Fig. 1 we see two typical examples of rain-induced image distortion within an automotive setting. Fig. 1A shows a typical example of ill-defined droplets that are barely visible within the overall context of the image but still effectively occlude the localised image region, causing significant distortion. Such examples are difficult to detect due to low contrast differences with their surroundings and the lack of any clearly defined boundaries. By contrast, Fig. 1B shows well-defined,

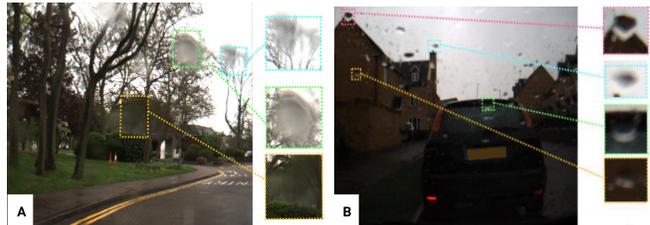
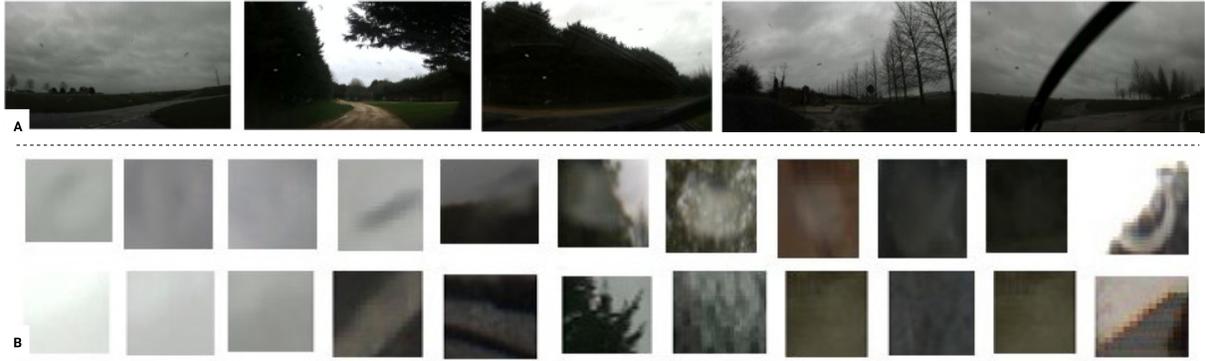


Fig. 1. Raindrop variation in terms of shape and colour [26].

stereotypical raindrops which are prominent within the image context. These examples are characterised by inverted scene illumination (dark top / light bottom) and exhibit clearly defined boundaries. In addition, within Figure 2A we can further appreciate the illumination challenges encountered, using standard automotive sensing cameras, under the general environment illumination conditions associated with significant rainfall. From this variation (Fig. 1 & 2) we can readily appreciate the broad challenge of automated raindrop detection and removal from later stages of any image understanding pipeline.

Prior work on explicit raindrop detection generally follows two themes:- (a) understanding of the photometric properties of raindrops [22] and (b) feature based detection and removal [31, 26]. The basis for photometric raindrop understanding leads to the ideal raindrop model [22]. Potential raindrop regions of a video image can thus be compared to the model forming a viable detection methodology [9]. Notably, variation within raindrops (e.g. Fig. 1) determines that not all such instances comply with any such model, compromising effective detection [22]. Furthermore, such approaches are characterised by high computational requirements limiting the applicability to real-time visual sensing applications [22, 9]. The second strand of raindrop detection work focuses on a range of feature-driven detection approaches [27, 29, 19, 28, 8]. The work of [30] uses stereo vision for raindrop detection whereby raindrops are detected via stereo matching, however atypical raindrops or poor conditions can cause failure. Later work [29], using a spatial-temporal approach, utilises consecutive video frames to analyse raindrops from a panning camera model in a visual surveillance context. However, this approach [29] requires prior knowledge of the camera dynamics and a relatively motion free scene between panning operations.

Within the automotive context, the work of [8] investigates windshield raindrop detection with the use of a specific hardware set up with the aim to detect raindrops through a reflection and focusing principle. A lens arrangement brings



**Fig. 2.** Illumination variations within automotive scenes (A) and isolated raindrop regions (B).

the raindrops into focus and blurring the background with a bespoke LED arrangement providing necessary illumination for nocturnal use. Raindrops are then detected by distinguishing sharp, in-focus regions in the image [8] but only over a limited field of view. Also motivated by automotive sensing, the work of Wu et al. [27] investigates the idea that raindrops are salient regions in the field of view and puts forward an adapted saliency map for detecting these regions within the XYZ colour space. A multi-scale approach derives colour, texture saliency and rudimentary shape information as feature maps over all three colour channels (for first stage candidate detection). These form an input to Support Vector Machine (SVM) [27] classification (for second stage classification). Although [27] produces strong detection results (precision / recall = 0.79 / 0.76, outperforming [19, 22]) it has been found to suffer from notably high false positives (Table 1).

The automotive focused work of [26] extended the saliency method proposed by [27] with a specific focus on improving both recall and precision, by adopting an extended feature descriptor comprising localised Hu shape moments and using saliency / texture features that are first isolated from the overall scene context. Furthermore, the work of [26] also considered Random Forest (RF) classification as a comparator to the earlier SVM classification approach of [27] (precision / recall = 0.86 / 0.92 outperforming [27] - Table 1).

This earlier work [27, 26] relied on an initial first stage of feature-based candidate detection, via the use of saliency / texture based feature maps, from which a second stage of SVM/RF classification operates on a hand-tuned feature-space encoded via classical bag of visual words encoding. By contrast today we would consider the use of a contemporary Convolutional Neural Network (CNN) architecture [20], that can in its earlier form consider whole-image classification [18] or more recently end-to-end region-based detection (object localisation) and classification [7]. Whilst the general direction of travel in efficient object detection is towards end-to-end CNN architectures [7, 21], comprising a jointly optimised region proposal network (framework) followed by a secondary classification network (akin to [18]) applied per region, translucent objects such as raindrops pose a particular challenging case. The joint optimisation of architectures such

as [21], and similar inter-changeable variants [1], rely on the concept of well defined objectness in order to learn how to detect regions likely to contain the set of target objects - as illustrated within Figures 2 and 1, this concept is significantly lacking for such translucent and variable objects making such architectures largely redundant. This forces the use of consideration of explicit generalised region proposal strategies as a forerunner to any CNN classification approach.

In this work, we consider the impact of varying these region proposal strategies for first stage (exhaustive) candidate detection from which we show that strong second stage candidate classification (discrimination), via novel down-scaled variants of seminal CNN architectures, is fully capable of delivering an overall raindrop detection approach that comparatively outperforms the current state of the art [27, 26].

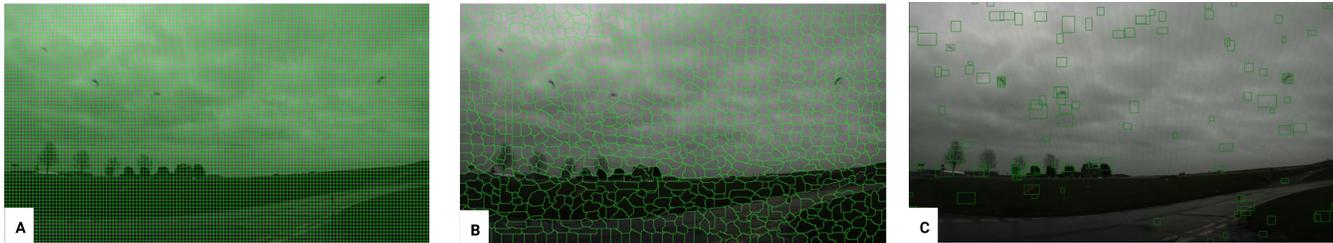
## 2. RAINDROP DETECTION

Our approach considers three region proposal methodologies (Section 2.1) over which we propose raindrop scale-specific variants to established deep CNN architectures for final region classification (Section 2.2).

### 2.1. Region Proposal

The first region proposal methodology considered is that of a simple sliding window of dimension  $w \times h$  and overlap stride  $s$ . As illustrated in Figure 3A, this generates a significant number of candidate regions for secondary processing by the CNN classification model.

Our second methodology considers the use of image over-segmentation in the form of superpixels - specifically the highly efficient SEEDS work of [25]. Superpixels are perceptually meaningful as each represents a consistent unit of colour, texture and shape and therefore the pixels within each superpixel boundary can be considered locally consistent to even small perturbations across aspects of sharpness, colour reflection or contours associated with raindrop structures (Fig. 1). SEEDS starts from a complete superpixel partitioning at a specified resolution,  $w_{SEEDS}$ , which is then refined over  $i_{SEEDS}$  iterations. This refinement is carried out, via a hill-climbing energy optimisation approach, such that superpixel boundaries conform to the localised structural boundaries within the image optimising the objective function



**Fig. 3.** Exemplar region proposals generated by each of the proposal methodologies.

in use. This results in a set of candidate regions conforming to local shape boundaries as illustrated in Fig. 3B.

Finally, we consider the selective search methodology of [24], which addresses the problem of generating possible object locations in the scene based on a combined concept which leverages the strength of both exhaustive search and segmentation. Segmentation is used to guide the location sampling process whilst an exhaustive element aims to capture all possible object locations supporting various partitionings to facilitate numerous possible image conditions. Selective search results in a smaller set of high-quality multi-scale class-independent candidate regions within the image [24] (using  $scale = minimum\ size = 50$ ; Gaussian  $\sigma = 0.7$  from [24]). Here the resulting set of candidate regions are post-filtered based on size and height ( $h$ ) to width ( $w$ ) ratio ( $h : w < 1.5$ ;  $w : h < 1.5$ ) to isolate a set of raindrop scale candidates as illustrated in Fig. 3C.

## 2.2. Region Classification

Each candidate region is subsequently classified using a deep CNN architecture model [20]. Three network architectures are considered to form the base of our novel scale-specific architectures. The seminal AlexNet architecture [18], the InceptionV1 modular architecture of [23] (GoogLeNet) and an AlexNet-Lite ([18] with each layer first reduced in size by 50%) are all re-formulated to use a down-scaled  $30 \times 30$  (raindrop size) input layer. This requires redefinition to the subsequent convolution, pooling and fully-connected layers of these architectures, with automatic padding as needed to prevent layer collapse, following the relative layer definitions of the original architectural model (now denoted as AlexNet-30<sup>2</sup>, InceptionV1-30<sup>2</sup> and AlexNet-Lite-30<sup>2</sup>). In each case, this results in a significantly smaller network size, suitable for the scale of input image within the raindrop classification task. The InceptionV1 architecture offers a contrasting 22 layer deep network architecture to AlexNet (8 layers), offering superior benchmark performance [23], whilst having 12 times fewer parameters through modularisation. All three network variants are trained from random initialisation using stochastic gradient descent with a momentum of 0.9 and categorical cross-entropy loss. The initial learning rate of 0.001 is used with a decay rate of 0.96 applied every 100 iterations over a batch size of 64.

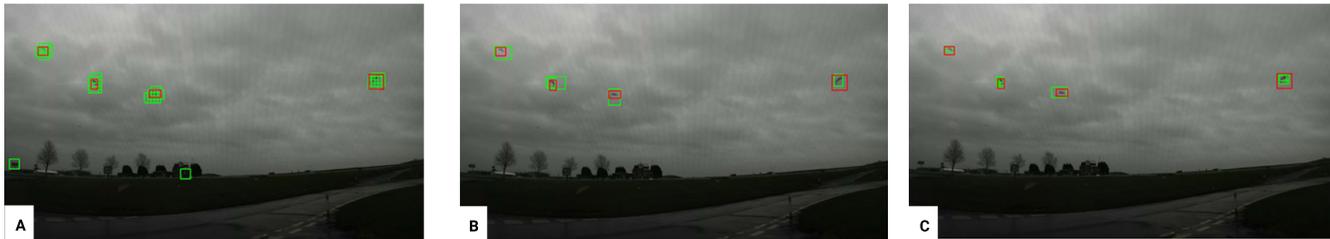
In addition to our re-scaled versions of these seminal network architectures, we also consider comparison against

contemporary approaches of SqueezeNet [15] offering equivalent accuracy to the AlexNet architecture with  $50\times$  fewer parameters and a  $< 0.5Mb$  model size (SqueezeNet-1.0) or alternatively  $2.4\times$  less computation and slightly fewer parameters again without sacrificing accuracy (SqueezeNet-1.1). Furthermore, we consider additional comparison against the more recent high-performance, large-scale ResNet [12] and DenseNet CNN architectures [14] over varying network depths,  $D$  (i.e. number of layers, ResNet- $D$  / DenseNet- $D$ ). By contrast, the latter network variants (SqueezeNet, ResNet and DenseNet) are trained based on their original network architecture with up-scaling of the training samples to the original input size of the network. Full architectural details are presented within the original works [18, 23, 15, 12, 14]. All networks are trained using a Nvidia GeForce 840M (AlexNet-30<sup>2</sup>, InceptionV1-30<sup>2</sup>, AlexNet-Lite-30<sup>2</sup>) via TensorFlow (1.1 + TFLearn 0.3) or a Nvidia Titan X GPU (all others) via Pytorch (0.3).

## 3. EVALUATION

For the comparison of the region proposal and CNN architectures outlined we consider True Positives (TP), True Negatives (TN), False Positives (FP), False Negatives (FN) together with the Precision (P) Recall (R) and accuracy (A) statistics against the recent work of Webster et al. [26] and two variants of Wu et al. [27] (as reported in [26]; Table 1).

Classifier training was performed in two stages as set out in Section 2.2. Each network was first trained on the dataset available from Webster et al. [26] (4365 raindrop and 4401 non-raindrop examples) and tested over the separate independent test set from the same (4554 raindrop and 4014 non-raindrop examples) extracted from  $1024 \times 768$  resolution source video. A second stage of training was then performed over four separate datasets (1000 raindrops / 1000 non-raindrops) each constructed using both the Webster et al. [26] examples and additional image sets specifically targeting regions containing of tree, sky, road and general (windshield wiper, building, grass etc.) extracted from  $1280 \times 720$  resolution source video. Additional data augmentation was performed using vertical and horizontal image flipping and rotation by  $\pm 45^\circ$  on the positive (raindrop) examples only. Whilst the first stage provides good performance on light spot distractions, the second stage improves performance over cloud, road and tree branch regions. Final validation testing, as reported in Table 1, was performed over a separate data set consisting of 399 non-raindrop and 400 raindrop examples.



**Fig. 4.** Example detection results via of the proposed methodologies [*Green = detected; RED = ground truth*].

From the results presented in Table 1, we can see that our CNN variants (AlexNet-30<sup>2</sup>, InceptionV1-30<sup>2</sup> and AlexNet-Lite-30<sup>2</sup>) significantly outperform prior work of [26] and [27] (Table 1, upper section). Furthermore, AlexNet-30<sup>2</sup> offers the lowest overall FP, highest TN, maximal precision and accuracy with 1% of the minimum achieved across all techniques evaluated. AlexNet-Lite-30<sup>2</sup> performs marginally worse than the AlexNet-30<sup>2</sup> architecture across the board with notably higher FP but lower FN. InceptionV1-30<sup>2</sup> performs within 1% of AlexNet-30<sup>2</sup> across precision, recall and accuracy. Furthermore, our novel re-scaled CNN architectures outperform the competing reduced complexity SqueezeNet architecture [15] (Table 1, upper + middle section). Indeed, only the significant complexity of the ResNet/DenseNet CNN architecture marginally outperform our low-complexity approaches (Table 1, lower; for  $D > 50$ ).

Performance evaluation of the three region proposal approaches outlined was performed over a set of 240 images (1280×720 resolution) annotated with in-frame ground truth rain drop locations (1868 ground truth raindrops) gathered using a forward facing digital camera mounted behind a car windscreen under a variety of road environments and weather (rainy) conditions (e.g. Fig. 2). For all three region proposal variants, the maximally performing AlexNet-Lite-30<sup>2</sup> architecture (lowest FP, highest P, within top 2% of accuracy) was used for the secondary stage of region classification. Dependent on the region proposal approach in use, this generated approximately 550-700 regions per image that intersected with the ground truth raindrop examples (true positives) and a varying set of non-raindrop regions depending on the technique (maximal true negatives: 9400 for sliding window, 1000 for SEEDS, 800 for selective search). This characterises the high TN and very low FP rates shown (Table 1).

From the results presented in Table 1 (lower section), we can see that region proposal approach in use has a significant impact on in-frame raindrop classification performance (illustrated in Fig. 4). Whilst the computationally expensive sliding window approach provides maximal overall accuracy, it is notable that its maximal performance is comparable to the superpixel based SEEDS [25] approach (for  $W_{SEEDS} = 1000$ ;  $i_{SEEDS} = 12$ ) to within 2-3% across all evaluation measures. By contrast use of selective search region proposal [24] produces significantly worse overall performance (Table 1, lower section). Combined with the novel proposition of the AlexNet-30<sup>2</sup> architecture for secondary classification, we

see overall in-frame detection accuracy of 0.94 significantly outperforming the state of the art [27, 26] (Table 1).

	TP%	TN%	FP%	FN%	P	R	A
Wu et al. [27] (RF)	60.24	77.93	22.07	34.91	0.73	0.65	0.72
Wu et al. [27] (SVM)	79.41	13.24	86.76	14.19	0.48	0.86	0.48
Webster et al. [26]	79.17	93.15	6.85	14.46	0.92	0.86	0.86
<b>AlexNet-30<sup>2</sup></b>	92.69	<b>96.96</b>	<b>3.03</b>	7.31	<b>0.97</b>	0.93	<b>0.95</b>
AlexNet-Lite-30 <sup>2</sup>	91.41	96.13	9.20	3.87	0.91	0.86	0.94
InceptionV1-30 <sup>2</sup>	96.27	90.63	4.46	7.90	0.96	0.92	0.94
SqueezeNet-1.0	92.23	89.09	10.91	7.77	0.91	0.92	0.91
SqueezeNet-1.1	94.01	87.29	12.71	5.99	0.89	0.94	0.91
ResNet-18	94.27	95.14	4.86	5.73	0.96	0.94	0.95
ResNet-50	96.84	89.66	10.34	3.16	0.91	<b>0.97</b>	0.93
ResNet-101	96.22	91.58	8.42	3.78	0.93	0.96	0.94
ResNet-152	96.14	95.69	4.31	3.86	0.96	0.96	<b>0.96</b>
DenseNet-121	96.93	95.84	4.16	3.07	0.96	<b>0.97</b>	<b>0.96</b>
DenseNet-161	96.75	93.30	6.70	3.25	0.94	<b>0.97</b>	0.95
<b>DenseNet-201</b>	<b>97.28</b>	95.42	4.58	<b>2.72</b>	0.96	<b>0.97</b>	<b>0.96</b>
Sliding W. ( $h = w = 30$ )	<b>91.29</b>	<b>99.99</b>	<b>0.01</b>	<b>8.71</b>	0.55	<b>0.91</b>	<b>0.95</b>
<b>SEEDS</b> ( $w = 1000; i = 12$ )	<b>88.65</b>	<b>99.94</b>	<b>0.06</b>	<b>11.34</b>	0.57	<b>0.89</b>	<b>0.94</b>
SEEDS ( $w = 1000; i = 20$ )	83.48	99.73	0.60	16.42	<b>0.69</b>	0.83	0.91
Selective Search	71.29	97.84	2.16	28.71	0.61	0.71	0.85

**Table 1.** Comparative statistical results

Figure 4 shows exemplar raindrop detection using each of the sliding window (Fig. 4A), SEEDS (Fig. 4B) and selective search (Fig. 4C) approaches, where we see detected regions classified as raindrops (green border) and additionally the ground truth raindrop locations for the image (red border).

#### 4. CONCLUSION

This work explores varying region proposal strategies for in-frame raindrop detection within the context of using novel, down-scaled seminal CNN architectures for secondary classification. Within this context, a superpixel based region proposal strategy is shown to deliver overall maximal performance (accuracy: 0.94, outperforming prior work of [27, 26]) based on the use of a novel AlexNet-30<sup>2</sup> architectural variant for the classification stage. Comparison of a set of novel, down-scaled versions of seminal CNN architectures (AlexNet-30<sup>2</sup>, InceptionV1-30<sup>2</sup> and AlexNet-Lite-30<sup>2</sup>) based around the concept of small (raindrop) size image region inputs show superior performance against competing low-complexity CNN architectures and comparable performance to leading high-complexity architectures for this task.

## 5. REFERENCES

- [1] S. Akcay and T.P. Breckon. An evaluation of region based object detection strategies within x-ray baggage security imagery. In *Proc. Int. Conf. on Image Processing*. IEEE, September 2017.
- [2] A. Atapour-Abarghouei and T.P. Breckon. Depthcomp: Real-time depth image completion based on prior semantic scene segmentation. In *Proc. British Machine Vision Conf.* BMVA, September 2017.
- [3] T.P. Breckon, A. Gaszczak, J. Han, M.L. Eichner, and S.E. Barnes. Multi-modal target detection for autonomous wide area search and surveillance. In *Proc. SPIE Emerging Technologies in Security and Defence: Unmanned Sensor Systems*, volume 8899, pages 1–19. SPIE, September 2013.
- [4] M. Breszcz and T.P. Breckon. Real-time construction and visualization of drift-free video mosaics from unconstrained camera motion. *IET J. Engineering*, 2015(16):1–12, August 2015.
- [5] M. L. Eichner and T.P. Breckon. Integrated speed limit detection and recognition from real-time video. In *Proc. IEEE Intelligent Vehicles Symposium*, pages 626–631. IEEE, June 2008.
- [6] K. Garg and S.K. Nayar. Detection and removal of rain from videos. In *Proc. Conf. Computer Vision and Pattern Recognition*, volume 1, pages I–I. IEEE, 2004.
- [7] R. Girshick, J. Donahue, T. Darrell, and J. Malik. Rich feature hierarchies for accurate object detection and semantic segmentation. In *Proc. Conf. on Computer Vision and Pattern Recognition*, pages 580–587, 2014.
- [8] S. Gormer, A. Kummert, S.B. Park, and P. Egbert. Vision-based rain sensing with an in-vehicle camera. In *Proc. Intelligent Vehicles Symp.*, pages 279–284. IEEE, 2009.
- [9] J.C. Halimeh and M. Roser. Raindrop detection on car windshields using geometric-photometric environment construction and intensity-based correlation. In *Proc. Intelligent Vehicles Symp.*, pages 610–615. IEEE, 2009.
- [10] O.K. Hamilton and T.P. Breckon. Generalized dynamic object removal for dense stereo vision based scene mapping using synthesised optical flow. In *Proc. Int. Conf. on Image Processing*, pages 3439–3443. IEEE, September 2016.
- [11] O.K. Hamilton, T.P. Breckon, X. Bai, and S. Kamata. A foreground object based quantitative assessment of dense stereo approaches for use in automotive environments. In *Proc. Int. Conf. on Image Processing*, pages 418–422. IEEE, September 2013.
- [12] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In *Proc. Conf. Computer Vision and Pattern Recognition*, pages 770–778, 2016.
- [13] C.J. Holder, T.P. Breckon, and X. Wei. From on-road to off: Transfer learning within a deep convolutional neural network for segmentation and classification of off-road scenes. In *Proc. European Conf. on Computer Vision Workshops*, pages 149–162. Springer, October 2016.
- [14] G. Huang, Z. Liu, K. Q. Weinberger, and L. van der Maaten. Densely connected convolutional networks. In *Proc. Conf. Computer Vision and Pattern Recognition*, volume 1, page 3, 2017.
- [15] F.N. Iandola, S Han, M.W. Moskewicz, K. Ashraf, W.J. Dally, and K. Keutzer. Squeezenet: Alexnet-level accuracy with 50x fewer parameters and < 0.5 mb model size. *arXiv preprint arXiv:1602.07360*, 2016.
- [16] I. Katramados and T.P. Breckon. Dense gradient-based features (degraf) for computationally efficient and invariant feature extraction in real-time applications. In *Proc. Int. Conf. on Image Processing*, pages 300–304. IEEE, September 2016.
- [17] T. Kriechbaumer, K. Blackburn, T.P. Breckon, O. Hamilton, and M. Riva-Casado. Quantitative evaluation of stereo visual odometry for autonomous vessel localisation in inland waterway sensing applications. *Sensors*, 15(12):31869–31887, December 2015.
- [18] A. Krizhevsky, I. Sutskever, and G.E. Hinton. ImageNet Classification with Deep Convolutional Neural Networks. In *Advances in Neural Information Processing Systems*, pages 1097–1105, 2012.
- [19] H. Kurihata, T. Takahashi, I. Ide, Y. Mekada, H. Murase, Y. Tamatsu, and T. Miyahara. Rainy weather recognition from in-vehicle camera images for driver assistance. In *Proc. Intelligent Vehicles Symp.*, pages 205–210. IEEE, 2005.
- [20] Y. LeCun, Y. Bengio, and G. Hinton. Deep learning. *Nature*, 521(7553):436, 2015.
- [21] S. Ren, K. He, R. Girshick, and J. Sun. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pages 1–1, Jun 2016.
- [22] M. Roser and A. Geiger. Video-based raindrop detection for improved image registration. In *Proc. Int. Conf. Computer Vision Workshops*, pages 570–577. IEEE, 2009.
- [23] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich. Going deeper with convolutions. In *Proc. Conf. Computer Vision and Pattern Recognition*, pages 1–9, 2015.
- [24] J. R. R. Uijlings, K. E. A. van de Sande, T. Gevers, and A. W. M. Smeulders. Selective search for object recognition. *Int. J. of Computer Vision*, 104(2):154–171, 2013.
- [25] M. Van den Bergh, X. Boix, G. Roig, B. de Capitani, and L. Van Gool. Seeds: Superpixels extracted via energy-driven sampling. In *Proc. Int. Conf. on Computer Vision*, pages 13–26. Springer, 2012.
- [26] D.D. Webster and T.P. Breckon. Improved raindrop detection using combined shape and saliency descriptors with scene context isolation. In *Proc. Int. Conf. on Image Processing*, pages 4376–4380. IEEE, September 2015.
- [27] Q. Wu, W. Zhang, and B V. Kumar. Raindrop detection and removal using salient visual features. In *Proc. Int. Conf. Image Processing (ICIP)*, pages 941–944. IEEE, 2012.
- [28] A. Yamashita, I. Fukuchi, T. Kaneko, and K.T. Miura. Removal of adherent noises from image sequences by spatio-temporal image processing. In *Proc. Int. Conf. Robotics and Automation*, pages 2386–2391. IEEE, 2008.
- [29] A. Yamashita, T. Harada, T. Kaneko, and K.T. Miura. Removal of adherent noises from images of dynamic scenes by using a pan-tilt camera. In *Proc. Int. Conf. Intelligent Robots and Systems*, volume 1, pages 437–442. IEEE, 2004.
- [30] A. Yamashita, M. Kuramoto, T. Kaneko, and K.T. Miura. A virtual wiper-restoration of deteriorated images by using multiple cameras. In *Proc. Int. Conf. Intelligent Robots and Systems*, volume 4, pages 3126–3131. IEEE, 2003.
- [31] S. You, R.T. Tan, R. Kawakami, and K. Ikeuchi. Adherent raindrop detection and removal in video. In *Proc. Conf. Computer Vision and Pattern Recognition*, pages 1035–1042. IEEE, 2013.