On the Impact of Illumination-Invariant Image Pre-transformation for Contemporary Automotive Semantic Scene Understanding

Naif Alshammari[1,2], Samet Akcay[1], and Toby P. Breckon[1]

Abstract— **Illumination changes in outdoor environments under non-ideal weather conditions have a negative impact on automotive scene understanding and segmentation performance. In this paper, we present an evaluation of illumination-invariant image transforms applied to this application domain. We compare four recent transforms for illumination invariant image representation, individually and with colour hybrid images, to show that despite assumptions to contrary such invariant pre-processing can improve the state of the art in scene understanding performance. In addition, we propose a robust approach based on using an illumination-invariant image representation, combined with the chromatic component of a perceptual colour-space to improve contemporary automotive scene understanding and segmentation. By using an illumination invariant pre-process, to reduce the impact of environmental illumination changes, we show that the performance of deep convolutional neural network based scene understanding and segmentation can yet be further improved. This illuminating result enforces the need for invariant (unbiased) training sets within such deep network training and shows that even a well-trained network may still not offer truly optimal performance (if we ignore any prior data transforms attributable to *a priori* insight). Our approach is demonstrated over a range of example imagery where we show a notable improvement in performance using pre-processed, illumination invariant, automotive scene imagery.**

I. INTRODUCTION

Automotive scene understanding and segmentation is an active research topic requiring robust full scene understanding, under uncontrolled and varying illumination conditions. Understanding the entire scene for tasks such as image classification and segmentation for automotive applications is a very challenging task, made significantly more challenging by exposure to illumination conditions and weather changes in outdoor environments. Within this context, despite the success of recent deep convolutional neural network (CNN) techniques to achieve state-of-the-art performance for generalized automotive scene understanding [1]–[4], extreme illumination variation can lead to inaccurate scene classification and segmentation [5]–[10]. To attempt to overcome the impact of illumination conditions and isolate the outdoor scenes from these effects, we evaluate the use of illumination invariant image pre-processing as proposed in [5], [6], [7], [9] to reduce illumination variations and improve object classification and segmentation for automotive scenes.

To improve upon the recent advances in scene classification and segmentation, here we look to expand, explore and evaluate the use of illumination invariant transformations
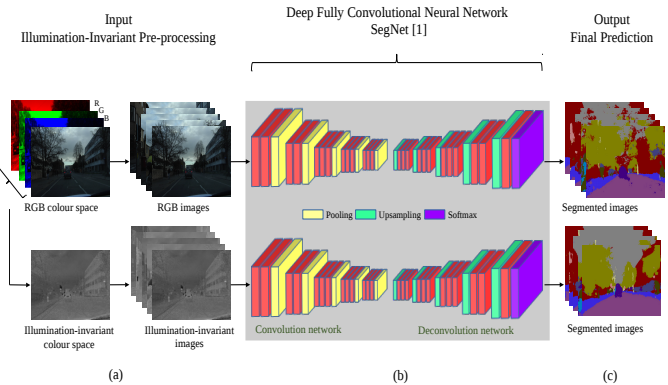
[1]Department of Computer Science, Durham University, UK
[2]Department of Natural and Applied Science, Community College, Majmaah University, Majmaah 11952, Saudi Arabia

Fig. 1: Our evaluation framework using two types of inputs: RGB and illumination invariant colour space (with and without hybrid colour information, $I_{\mathcal{I}HS}$) (a) and relying on deep fully convolutional neural network (CNN) architecture, SegNet [1] (b), as a model for object classification and semantic pixel-wise segmentation (c).

by assessing their impact as an initial pre-process prior to use of an existing fully deep convolutional neural network (CNN) architecture [1] for contemporary automotive scene understanding and semantic segmentation. Figure 1 shows our evaluation framework based on using an illumination-invariant colour space. When using a hybrid illumination invariant technique as a pre-process we find a significant improvement for classification within individual class performance and in overall semantic scene segmentation accuracy.

II. ILLUMINATION INVARIANT COLOUR SPACE

An illumination invariant colour space is a colour representation computed from RGB that removes (or minimises) scene colour variations due to varying scene lighting conditions. This technique was introduced as an intrinsic image to represent the illumination invariant and intrinsic properties in the image [11] with illumination transforms generally computed with reference to the physical properties behind the capture and presence of colour within the scene.

The use of illumination invariant approaches in most of the literature is predominantly for shadow removal [5], [12], [13], [8], [14], and to improve scene classification and segmentation [7], [6], [8]. Figure 2 shows an example of an RGB image (A) from [15] followed by four different illumination invariant images (B), (C), (D) and (E) generated using the approaches of [7], [6], [9], and [5]. In Section A we review recent approaches for illumination invariant image representation. Section B introduces the illumination-
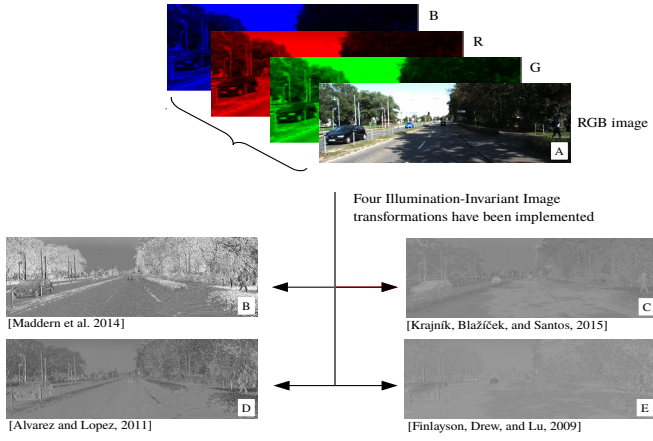
Fig. 2: An example of an RGB image (A) from KITTI dataset [15] followed by four different illumination invariant images, (B) [7], (C) [6], (D) [9], and (E) [5] where all the illumination variations such as shadows are significantly reduced within the scenes.

invariant image representation, combined with the chromatic components of a perceptual colour-space HSV [16].

A. Illumination Invariant Image Formulation

An illumination-invariant image $\mathcal{I}$ is a single channel image calculated by combining the three RGB colour channels in the image $I_{RGB} \in \{I_R, I_G, I_B\}$. For shadow removal purposes, [5] propose an approach to compute illumination invariant image by considering ρ_k as the response of a single RGB image sensor, $k \in \{I_R, I_G, I_B\}$ colour. Therefore, the illumination invariant images can be computed as follows:

$$\rho_k = \sigma \int E(\lambda)S(\lambda)R_k(\lambda)d\lambda, \qquad (1)$$

where $E(\lambda)$ is the illumination of the surface, $S(\lambda)$ is the surface reflectance, $R_k(\lambda)$ is the camera sensitivity of RGB colour sensors, and λ is the wave length. Equation 1 can be simplified to:

$$\rho_k = \sigma E(\lambda_k)S(\lambda_k)q_k, \qquad (2)$$

where q_k is Planckian illuminant and infinitely narrow-band color sensitivities.

The work of [5] approximate lighting by using Planck's law, which utilizes the Wien approximation by colour temperature T as follows:

$$E(\lambda, T) = I_{c1}\lambda^{-5}e^{-\frac{c_2}{T\lambda}}, \qquad (3)$$

where c_1, c_2 are constant and I characterizes the lighting colour.

From Equation 3, the RGB colour is simply given by

$$\rho_k = \sigma I c_1 \lambda_k^{-5} e^{-\frac{c_2}{T\lambda}} S(\lambda_k)_{qk}. \qquad (4)$$

Now the band-ratio 2-vector chromaticities c is calculated as follows:

$$c_k = \rho_k/\rho_p, \qquad (5)$$

where p is one of the channels, and k is the remaining responses.

A straight line parametrized by T, can be obtained as follows:

$$\rho_k \equiv \log(c_k) = \log(s_k/s_p) + (e_k - e_p)/T. \qquad (6)$$

A sample illumination image $\mathcal{I}_{Finlayson}$ (as shown in Fig. 2E) is thus computed by projecting 2D logs in the chromaticity space into a direction on the illumination invariant vector [5].

Inspired by [12] and focusing on one class problem, [7], [17] proposes an approach for road detection and segmentation relying on illumination invariant colour space, as introduced by [12], to reduce the complicity of the scenes when the shadows are present. Within [7] the illumination invariant image $\mathcal{I}$ (a sample shown in Fig. 2D) is compute by combining the RGB channels as follows:

$$\begin{aligned} \mathcal{I}_{\text{Álvarez}} = \ & \cos(\theta) \log_{approx}(I_R/I_B) + \\ & \sin(\theta) \log_{approx}(I_G/I_B), \end{aligned} \qquad (7)$$

where I_R, I_G, I_B are the three RGB channels, $\theta \in \{0...180\}$, and $\log_{approx}()$ is the logarithmic approximation for the division of two channels in an image $x \in \{R, G, B\}$, which is computed as follows:

$$\log_{approx(x)} = \alpha((\vec{x}^{\frac{1}{\alpha}}) - 1), \qquad (8)$$

where x is the value from dividing two channels, $\alpha = 5000$.

In the work of [6], an illumination-invariant approach is proposed to improve visual localization, mapping and scene classification for autonomous road vehicles. In the aforementioned approach, the illumination-invariant image $\mathcal{I}$ (as shown in Fig. 2B) is computed by converting a 3-channel floating point RGB image into the corresponding illumination invariant image as follows:

$$\begin{aligned} \mathcal{I}_{\text{Maddern}} = \ & 0.5 + \log(I_G) - \alpha log(I_B) - \\ & (1 - \alpha)\log(I_R), \end{aligned} \qquad (9)$$

where $\alpha = 0.48$ for the reference camera in use (PointGrey Bumblebee-2), and 0.5 for pixels normalised into the range $\{0...1\}$. Unlike [5] and [7], the work of [6] addresses the problem of multiple classes (e.g. buildings, road, vehicles, vegetation and sky) for detection and semantic segmentation tasks.

In order to improve mobile robot localization in outdoor environments, [9] calculates the illumination-invariant image $\mathcal{I}$ (see Fig. 2C) by projecting 3D log-chromaticity to a 2D log-chromaticity space $\psi_{1,2}$ using a 3D to 2D space transformation matrix U. Subsequently, knowing the suitable estimated angle θ estimated based on the minimization of Shannons entropy, the intrinsic image is computed as follows:

$$\mathcal{I}_{\text{Krajník}} = (\psi_1 \cos\theta + \psi_2 \sin\theta), \qquad (10)$$

where θ is a suitable estimated angle calculated based on the minimization of Shannon entropy of the intrinsic image.

The related work of [8] proposes a street classification and semantic segmentation approach based on using illumination invariant transforms, which are able to significantly reduce illumination variation in outdoor scenes. To compute the illumination invariant images, [8] use a single channel feature

space $\mathcal{I}$ combined with three linear sensors $\{R, G, B\}$ which is very similar to [6] ($\mathcal{I}_{\text{Maddern}}$) as follows:

$$\mathcal{I}_{\text{Upcroft}} = \log(I_G) - \alpha log(I_B) - (1 - \alpha) \log(I_R), \quad (11)$$

where again $\alpha = 0.48$ for the reference camera in use.

By contrast, here we evaluate the relative performance of each of these available set of invariant transforms, ($\mathcal{I}_{\text{Álvarez}}$[7], $\mathcal{I}_{\text{Maddern}}$[6], $\mathcal{I}_{\text{Krajník}}$[9], $\mathcal{I}_{\text{Finlayson}}$[5]), as a pre-transformation prior to segmentation and semantic labeling via a contemporary scene understanding methodology [1].
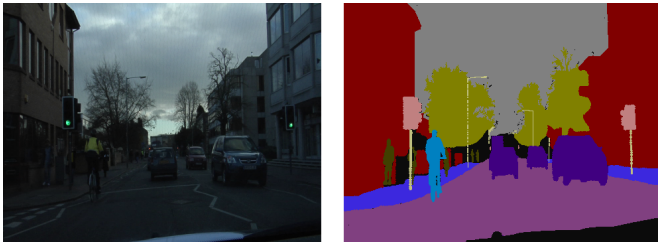

Fig. 3: An example image from the CamVid dataset [18] along with its annotations.

B. Perceptual Colour-Space and Illumination-Invariant

For further evaluation, we use the illumination invariant channel $\mathcal{I}$ proposed in [5], [6], [7], [9] combined with the chromatic components of a perceptual colour-space HSV. The only hue and saturation channels of an image $I_{HSV} \in \{I_H, I_S, I_V\}$ are combined with the illumination-invariant channel $\mathcal{I}$ to produce a new colour representation $I_{\mathcal{I}HS} \in \{I_{\mathcal{I}}, I_H, I_S\}$. We evaluate the use of this transformation on the performance of pixel-wise semantic scene segmentation.

III. EVALUATION METHODOLOGY

We evaluate the performance of automotive scene understanding and segmentation using the SegNet [1] CNN architecture (Figure 4) with four illumination-invariant transformations [6], [9], [7], [5]. Subsequently, we train the model again using the $I_{\mathcal{I}HS}$ formulation. We use the CamVid [18] dataset with different pixels classes for the SegNet classification task (Fig. 3 shows an example of the dataset along with its annotations). CamVid datasets have been used with eleven classes: {*sky, building, pole, road, pavement, tree, sign, fence, car, pedestrian, bicycle*}. This dataset consists of 367 training and 233 testing image examples (resolution of 480×360).

A. Network Training

For the segmentation task, we train the CNN model SegNet [1] (the SegNet architecture is visualised in Fig. 4) on the CamVid dataset [18] with three different colour transformations: (1) the original CamVid dataset only [18] (i.e all are RGB images as per SegNet-Basic [1]); (2) the generated illumination-invariant datasets (computed using the approaches: [7], [6], [9], [5], reviewed in Section 2.1); and (3) the colour transformation we propose, $I_{\mathcal{I}HS}$. We use a VGG16 [19] network pre-trained on the ImageNet [20] dataset, which is the encoder network within SegNet
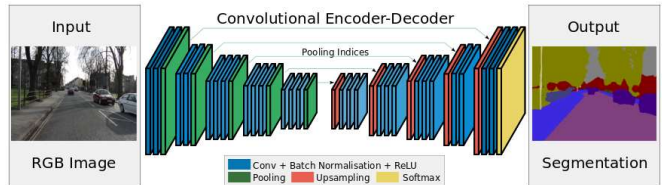

Fig. 4: Architecture of the SegNet Convolutional Neural Network [1].

[1]. An encoder network consists of convolution and pooling layers followed by a decoder network contains of convolution and upsampling layers. We employ a stochastic gradient descent (SGD) [21] optimisation with an initial learning rate of 1×10^{-3}, weight decay of 5×10^{-4} and momentum of 0.9. We train the model for 40,000 iterations by using an NVIDIA Tesla K40 GPU (with cuDNN v5) following the work of [1].

B. Analysis

To see the impact of the illumination invariant transformation on the original work of [1], we compare the semantic scene segmentation of the SegNet trained by using each of the illumination invariant transforms of [7], [6], [9], [5] and $I_{\mathcal{I}HS}$ with the original work of [1] using the CamVid [18] dataset. For semantic accuracy evaluation, we use the following evaluation measures: (1) class average accuracy which is the mean of the predictive accuracy over all classes, (2) global accuracy which measures overall scene pixel classification accuracy, and (3) mean intersection over union (mIoU). To supplement these measures, the following performance metrics have been used: (4) recall and (5) precision.

1) Illumination-invariant and RGB: We evaluate the performance of scene understanding and segmentation first on illumination-invariant transforms [7], [6], [9], [5] and the original transform (RGB) using images obtained from CamVid [18]. The reference results of [1] (shown in Table I and referred as Original (RGB)) produce the highest overall accuracy (80.76%), as well as when compared with the transforms: $\mathcal{I}_{\text{Álvarez}}$ (68.63%), $\mathcal{I}_{\text{Maddern}}$ (78.21%), $\mathcal{I}_{\text{Krajník}}$ (80.06%) and $\mathcal{I}_{\text{Finlayson}}$ (78.58%). However, the latter four illumination-invariant transforms excel in several sub-classes. For example, $\mathcal{I}_{\text{Krajník}}$ is distinct in each of the *Road* and *Bicyclist* classes, where their accuracy values are (93.6%) and (15.5%) respectively (Table I). In addition, $\mathcal{I}_{\text{Krajník}}$ has the second highest overall accuracy (80.06%). Accordingly, $\mathcal{I}_{\text{Maddern}}$ has the highest accuracy for the *Sky* class (92.4%), but its overall accuracy (78.21%) makes it the second least accurate transformation (Table I). Furthermore, $\mathcal{I}_{\text{Finlayson}}$ overall accuracy is (78.58%) (shown in Table I), placing it in the middle when compared with the other transform methods. However, $\mathcal{I}_{\text{Finlayson}}$ shows higher results in some bands, such as *Sky* and *Road*, especially when compared with the original results (with no pre-transform method). $\mathcal{I}_{\text{Álvarez}}$ has the lowest overall accuracy (68.63%), making it the worst performing method within the four illumination-invariant transforms discussed (see Table I).

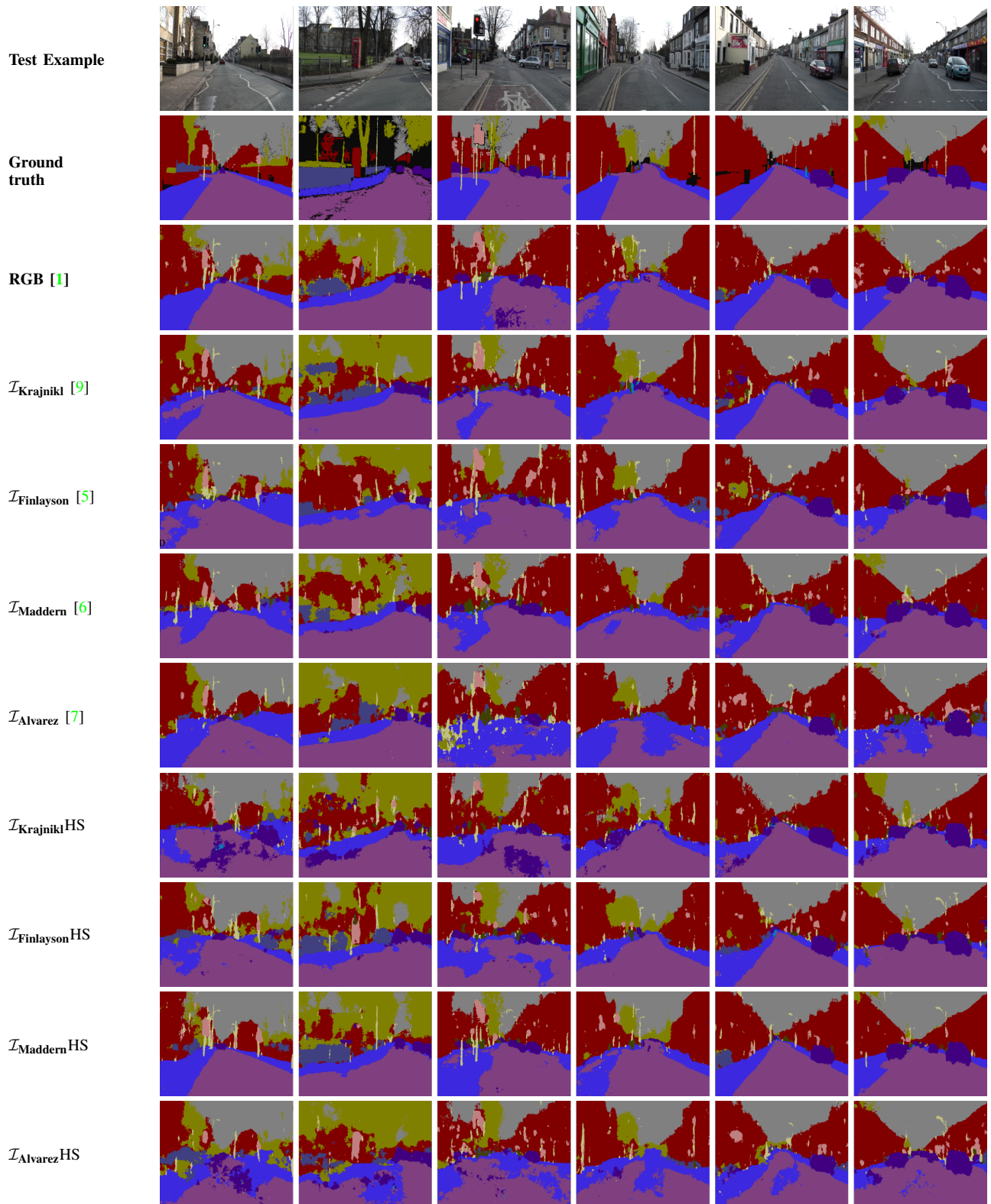|						
Test Example						
Ground truth						
RGB [1]						
$\mathcal{I}_{\text{Krajnikl}}$ [9]						
$\mathcal{I}_{\text{Finlayson}}$ [5]						
$\mathcal{I}_{\text{Maddern}}$ [6]						
$\mathcal{I}_{\text{Alvarez}}$ [7]						
$\mathcal{I}_{\text{Krajnikl}}$HS						
$\mathcal{I}_{\text{Finlayson}}$HS						
$\mathcal{I}_{\text{Maddern}}$HS						
$\mathcal{I}_{\text{Alvarez}}$HS						

Fig. 5: A comparison of classification predictions for SegNet [1] using RGB images obtained from [15], the illumination invariant transformations [9], [5], [6], [7], and $\mathcal{I}$HS for six scenarios.

Method	Sky	Building	Pole	Road	Pavement	Tree	SignSymbol	Fence	Car	Pedestrian	Bicyclist	Class avg.	Global acc.	mIoU	Precision	Recall
Original (RGB) [1]	0.73	0.846	0.33	0.87	**0.91**	0.76	0.43	0.41	0.73	0.60	0.11	0.61	0.807	0.46	**0.70**	0.61
$\mathcal{I}_{\text{Álvarez}}$ [7]	0.67	0.73	0.22	0.69	0.67	0.75	0.35	0.28	0.63	0.26	0.017	0.46	0.68	0.33	0.46	0.48
$\mathcal{I}_{\text{Maddern}}$ [6]	0.92	0.80	0.20	0.932	0.53	0.62	0.38	0.17	0.61	0.51	0.07	0.54	0.78	0.40	0.64	0.65
$\mathcal{I}_{\text{Finlayson}}$ [12]	0.88	0.77	0.21	0.92	0.63	0.67	0.37	0.27	0.68	0.36	0.05	0.53	0.78	0.41	0.53	0.53
$\mathcal{I}_{\text{Krajník}}$ [9]	0.88	0.805	0.28	0.936	0.61	0.70	0.42	0.30	0.70	0.40	0.15	0.56	0.806	0.44	0.56	0.57
$\mathcal{I}_{\text{Álvarez}}$HS	0.73	0.76	0.20	0.82	0.54	0.64	0.38	0.26	0.68	0.26	0.12	0.49	0.71	0.36	0.49	0.49
$\mathcal{I}_{\text{Maddern}}$HS	**0.93**	**0.87**	**0.37**	0.93	0.89	0.78	0.44	0.43	**0.79**	0.71	0.36	0.68	**0.873**	0.563	0.69	0.68
$\mathcal{I}_{\text{Finlayson}}$HS	0.91	0.81	0.307	**0.955**	0.67	0.69	**0.466**	0.28	0.75	0.43	0.30	0.61	0.81	0.58	0.58	0.60
$\mathcal{I}_{\text{Krajník}}$HS	0.92	0.86	0.39	0.94	0.86	**0.83**	0.55	**0.45**	0.78	**0.73**	**0.39**	**0.70**	0.872	**0.567**	0.68	**0.70**

TABLE I: Quantitative results are shown as accuracy of the CNN SegNet approach on CamVid test data [18] for three methods. First, the reference results of [1], followed by four illumination-invariant approaches of [7], [6], [5], [9], and our approach $\mathcal{I}$HS for the same set of transforms. The table shows the accuracy of predicted 11 classes, followed by class average, global accuracy, mIoU, precision, and recall.
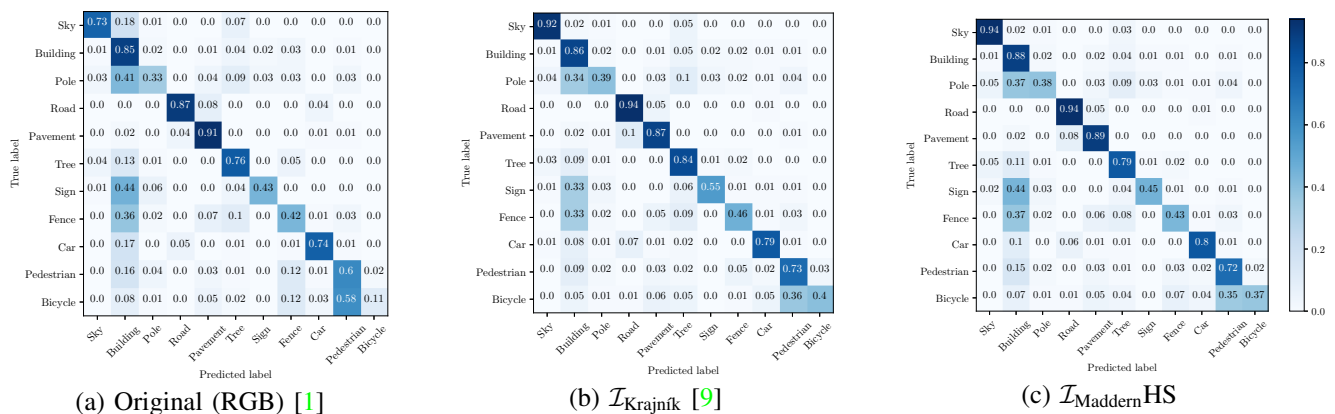

Fig. 6: Normalized confusion matrices based on the three transforms that gave the highest performance within their class categories (from Table I): (a) the reference results of [1] using RGB input, (b) the illumination-invariant method $\mathcal{I}_{\text{Krajník}}$ [9], and (c) our transformed method ($\mathcal{I}_{\text{Maddern}}$HS).

2) Illumination-invariant, RGB, and I$_{\mathcal{I}HS}$: The second evaluation is performed on datasets with the colour transformation $I_{\mathcal{I}\text{HS}}$. This approach achieved the highest overall segmentation accuracy (87.3%) specifically with ($\mathcal{I}_{\text{Maddern}}$HS) when compared with the aforementioned approaches (RGB and illumination-invariant) (see Table I). Furthermore, ($\mathcal{I}_{\text{Krajník}}$HS) (87.2%), and ($\mathcal{I}_{\text{Finlayson}}$HS) (81%) come the second and third highest overall accuracy. In addition to the results we gained in the overall accuracy, we attained the highest results in (class average, mIoU, Recall) (70%, 56%, 70%) respectively, using ($\mathcal{I}_{\text{Krajník}}$HS). However, the reference results of [1] excels in precision where it represents (70%). For per-class comparison, our proposed approach offers superior performance in ten out eleven classes. In particular, $\mathcal{I}_{\text{Maddern}}$HS has the highest accuracy in each of the {*Sky, Building, Pole, Car*} classes where they represent (93%, 87%, 95%, 79%) respectively whilst $\mathcal{I}_{\text{Finlayson}}$HS) is distinct in each of the {*Road, SignSymbol*} classes with accuracy values (95%, 46.6%) respectively. Lastly, $\mathcal{I}_{\text{Krajník}}$HS excels in each of the {*Tree, Fence, Pedestrian, Bicyclist*} classes where they represent (83%, 45%, 73%, 39%) respectively while the only distinction by the reference results of [1] within this context is in the class *Pavement* (91%). Figure 5

shows a comparison of classification predictions on CamVid dataset [18] with three different colour transformations:- (1) RGB (no transform method) (2) the illumination invariant transformations in [7], [6], [9], [5] (3) and our outlined approach ($I_{\mathcal{I}\text{HS}}$) computed using aforementioned illumination-invariant transformations for six scenarios. Figure 6 shows three normalized confusion matrices that depict per-class accuracy based on implementing the aforementioned transforms and representing the transformations that achieve the highest performance within their class categories (from Table I).

C. Discussion

In exploring the use of the illumination invariant technique for scene understanding and pixel-wise semantic segmentation, as compared with RGB, we find that illumination invariant transformations show significant improvement in prediction certain classes such as {*Road, Sky*}. However, the illumination-invariant techniques fail to achieve the overall accuracy (in our case where we are segmenting eleven classes). Most of the literature illumination-invariance within the context of scene understanding and semantic segmentation predominantly uses the illumination invariant technique

for a one class problem, *Road* in [7], [22], and [23], or for limited number of classes in [6], [8] (seven classes). By contrast, this study presents an extended evaluation over an extended set of classes and illumination invariant transforms in a side-by-side comparison.

On the other hand, the use of an illumination-invariant image representation, combined with the chromatic components of a perceptual colour-space HSV has improved robustness for scene understanding and semantic segmentation in both per-class and overall class accuracy using our proposed $I_{\mathcal{I}\text{HS}}$ transform (with $\mathcal{I}_{\text{Maddern}}$ [6], $\mathcal{I}_{\text{Finlayson}}$ [5], and $\mathcal{I}_{\text{Krajník}}$ [9]). Future work could additionally investigate the use of other perceptual colour spaces with invariant image transforms, following from the improved accuracy observed here.

Overall, we observe that the illumination invariant pre-transforms of [7], [6], [9], [5], and our outlined approach do impact the SegNet architecture performance either in the per-class accuracy, when using the illumination-invariant transformation ($\mathcal{I}$), or in the overall class accuracy, when using $I_{\mathcal{I}\text{HS}}$ transform (with $\mathcal{I}_{\text{Maddern}}$[6], $\mathcal{I}_{\text{Finlayson}}$ [5], and $\mathcal{I}_{\text{Krajník}}$ [9]). This reinforces the notation that even a well trained CNN approach can be improved by initial pre-processing against current conventional wisdom within the field on end-to-end semantic scene understanding architectures.

IV. Conclusion

In this paper, we present the impact of illumination-invariant image pre-transformation on contemporay automotive semantic scene understanding using deep convolutional neural networks (SegNet [1]). By examining the illumination invariant image transforms of [7], [6], [9], [5], and our outlined approach $I_{\mathcal{I}\text{HS}}$ as alternative inputs to SegNet based scene understanding and segmentation we are able to show that pre-processing can influence trained network performance. More notably, we achieve the highest (overall accuracy, class average, mIoU, recall) (87%, 70%, 56%, 70%) respectively for automotive scene segmentation by using our novel scene colour representations $I_{\mathcal{I}\text{HS}}$ with $\mathcal{I}_{\text{Maddern}}$HS giving the overall highest global accuracy. In addition, our approach demonstrates superior class accuracy in each of the {*Sky, Building, Pole, Road, Tree, SignSymbol, Fence, Car, Pedestrian, Bicyclist*}, (93%, 87%, 95%, 83%, 46.6%, 45%, 79%, 73%, 39%) respectively.

Hence, we are able to show that, contrary to conventional wisdom, the performance of deep convolutional neural network based scene understanding and segmentation can yet be further improved and influenced by input pre-processing. This shows that even a well trained network may still not offer truly optimal performance, and the impact of pre-processing in terms of dimensionality reduction and invariant feature representation may still offer performance gains. Ultimately, the internal feature representation within the network may not be truly optimal when compared to established *a priori* (recognised) insights on invariant scene understanding.

References

[1] V. Badrinarayanan, A. Kendall, and R. Cipolla, "Segnet: A deep convolutional encoder-decoder architecture for image segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 12, pp. 2481–2495, 2017. 1, 3, 4, 5, 6

[2] A. Krizhevsky, I. Sutskever, and G. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems 25*, 2012. 1

[3] 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 Cecognition*, 2014, pp. 580–587. 1

[4] S. C. Turaga, J. F. Murray, V. Jain, F. Roth, M. Helmstaedter, K. Briggman, W. Denk, and H. S. Seung, "Convolutional networks can learn to generate affinity graphs for image segmentation," *Neural Computation*, vol. 22, no. 2, pp. 511–538, 2010. 1

[5] G. Finlayson, M. Drew, and C. Lu, "Entropy minimization for shadow removal," *International Journal of Computer Vision*, vol. 85, no. 1, pp. 35–57, 2009. 1, 2, 3, 4, 5, 6

[6] W. Maddern, A. Stewart, C. McManus, B. Upcroft, W. Churchill, and P. Newman, "Illumination invariant imaging: Applications in robust vision-based localisation, mapping and classification for autonomous vehicles," in *Proc. Int. Conf. on Robotics and Automation*, vol. 2, 2014, p. 3. 1, 2, 3, 4, 5, 6

[7] J. Álvarez and A. Lopez, "Road detection based on illuminant invariance," *IEEE Trans. on Intelligent Transportation Systems*, vol. 12, no. 1, pp. 184–193, 2011. 1, 2, 3, 4, 5, 6

[8] B. Upcroft, C. McManus, W. Churchill, W. Maddern, and P. Newman, "Lighting invariant urban street classification," in *Proc. Int. Conf. on Robotics and Automation*, 2014, pp. 1712–1718. 1, 2, 6

[9] T. Krajník, J. Blažíček, and J. Santos, "Visual road following using intrinsic images," in *Proc. European Conference on Mobile Robots*, 2015, pp. 1–6. 1, 2, 3, 4, 5, 6

[10] P. Corke, R. Paul, W. Churchill, and P. Newman, "Dealing with shadows: Capturing intrinsic scene appearance for image-based outdoor localisation," in *Int. Conf on Intelligent Robots and Systems*, 2013, pp. 2085–2092. 1

[11] G. D. Finlayson and S. D. Hordley, "Color constancy at a pixel," *Journal of the Optical Society of America. A, Optics, Image Science, and Vision*, vol. 18, no. 2, pp. 253–264, 2001. 1

[12] G. Finlayson, S. Hordley, C. Lu, and M. Drew, "On the removal of shadows from images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 1, pp. 59–68, 2006. 1, 2, 5

[13] G. Finlayson, M. Drew, and C. Lu, "Invariant image improvement by srgb colour space sharpening," in *10th Congress of the International Colour Association.*, 2005, pp. 1–4. 1

[14] ——, "Intrinsic images by entropy minimization," in *European Conference on Computer Vision*. Springer, 2004, pp. 582–595. 1

[15] A. Geiger, P. Lenz, and R. Urtasun, "Are we ready for autonomous driving? the kitti vision benchmark suite," in *Proc. Conference on Computer Vision and Pattern Recognition*, 2012, pp. 3354–3361. 1, 2, 4

[16] C. Solomon and T. Breckon, *Fundamentals of Digital Image Processing*. John Wiley & Sons, Ltd, 2010. 2

[17] J. M. lvarez, A. M. Lpez, T. Gevers, and F. Lumbreras, "Combining priors, appearance, and context for road detection," *IEEE Transactions on Intelligent Transportation Systems*, vol. 15, no. 3, pp. 1168–1178, June 2014. 2

[18] G. Brostow, J. Fauqueur, and R. Cipolla, "Semantic object classes in video: A high-definition ground truth database," *Pattern Recognition Letters*, vol. 30, no. 2, pp. 88–97, 2009. 3, 5

[19] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *CoRR*, vol. abs/1409.1556, 2014. 3

[20] O. Russakovsky, J. Deng, H. Su, J. Krause, S. tobyatheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei, "ImageNet Large Scale Visual Recognition Challenge," *International Journal of Computer Vision*, vol. 115, no. 3, pp. 211–252, 2015. 3

[21] L. Bottou, "Large-scale machine learning with stochastic gradient descent," in *Proc. 19th Int. Conf. Comput. Statist.*, 2010, pp. 177–186. 3

[22] J. M. Álvarez, A. López, and R. Baldrich, "Illuminant-invariant model-based road segmentation," in *Proc. Intelligent Vehicles Symposium*, 2008, pp. 1175–1180. 6

[23] T. Kim, Y. Tai, and S. Yoon, "PCA based computation of illumination-invariant space for road detection," in *Proc. Winter Conference on Applications of Computer Vision*, 2017, pp. 1–6. 6