Outdoor Scenes Pixel-Wise Semantic Segmentation using Polarimetry and Fully Convolutional Network

Marc Blanchon, Olivier Morel, Yifei Zhang, Ralph Seulin, Nathan Crombez, Désiré Sidibé

Outdoor Scenes Pixel-Wise Semantic Segmentation using Polarimetry and Fully Convolutional Network

Marc Blanchon[1], Olivier Morel[1], Yifei Zhang[1], Ralph Seulin[1], Nathan Crombez[2] and Désiré Sidibé[1]
[1]*ImViA EA 7535, ERL VIBOT CNRS 6000, Université de Bourgogne Franche Comté (UBFC), 12 Rue de la Fonderie, 71200, Le Creusot, France*
[2]*EPAN Research Group, University of Technology of Belfort-Montbéliard (UTBM), 90010, Belfort, France*
marc_blanchon@etu.u-bourgogne.fr

Keywords: polarimetry, deep learning, segmentation, augmentation, reflective areas.

Abstract: In this paper, we propose a novel method for pixel-wise scene segmentation application using polarimetry. To address the difficulty of detecting highly reflective areas such as water and windows, we use the angle and degree of polarization of these areas, obtained by processing images from a polarimetric camera. A deep learning framework, based on encoder-decoder architecture, is used for the segmentation of regions of interest. Different methods of augmentation have been developed to obtain a sufficient amount of data, while preserving the physical properties of the polarimetric images. Moreover, we introduce a new dataset comprising both RGB and polarimetric images with manual ground truth annotations for seven different classes. Experimental results on this dataset, show that deep learning can benefit from polarimetry and obtain better segmentation results compared to RGB modality. In particular, we obtain an improvement of 38.35% and 22.92% in the accuracy for segmenting windows and cars respectively.

1 INTRODUCTION

Scene segmentation and understanding have been a popular topic in the field of robotics, artificial intelligence and computer vision. It has attracted a lot of research with different aproaches: decision forest approach (Gupta et al., 2014), deep approach for semantic segmentation (Couprie et al., 2013), and pixel-wise semantic segmentation (Badrinarayanan et al., 2015). The main challenge lies in the recognition and the assignment of multiple classes.
A difficult key point when addressing the problem of segmentation is the possible presence of reflective areas. The segmentation method should be able to differentiate a physical object and its projection on a reflective area.

The field of segmentation of complex scenes is open since many applications could benefit. Some research has been conducted on the detection of mud (Rankin and Matthies, 2010a), as well as on the detection of water (Yan, 2014; Nguyen et al., 2017). Indeed, robotics and autonomous cars could take advantage of these abilities. For example, if a system is able to understand a scene with complex areas (reflective), then it is possible to avoid them.

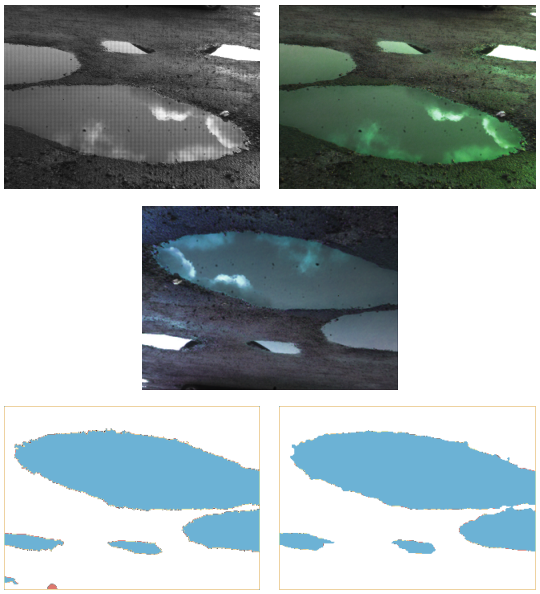

Figure 1: From raw polarimetric image to segmentation. Top: left is the raw polarimetric image, right is the transformed image to HSL (Hue Saturation Luminance). Middle image is the augmented image with proper physical meaning. Bottom: left is the hand made ground truth and right is the prediction of the deep learning network for the middle image.

To handle both the classification of so-called standard zones (or "low complexity") and areas of high complexity, the introduction of a discriminant modality is considered.
The choice is oriented towards the polarimetric imaging, giving the ability to measure and recover the changes in the light waves. SFP (shape-from-polarization) techniques have been using the ability of polarimetry to extract information from highly reflective objects (Rahmann and Canterakis, 2001; Morel et al., 2005). Therefore, polarimetric cameras have experienced a big development leading to better ease of use and practicality. The Division of Focal Plane (DoFP) allows the capture of an image using four different polarizers. In consequence, it is similar as acquiring four images with four polarizers.
Combining the advantages of different data types, a polarimetric camera will process non-reflective data as usual gray-scale portion of the image, while reflective areas will observe changes in the image information. In consequence of using polarimetric images, a set of constraints has been deduced to design a data augmentation process.
Since the aim of this paper is to measure and qualify the usefulness of a complex modality applied to a specific task, it is unnecessary to complexify the task at the early stage of the processing. Consequently, a widely used and tested network is the core of this study: SegNet (Badrinarayanan et al., 2015). The robustness and modularity of this architecture makes this network the perfect candidate for our purpose.
As shown in Figure 1, this paper allows understanding and exploitation of this new type of information in the context of deep learning.
This paper proposes the following main contributions:

- Introduction of the polarimetry in the field of feature learning to discuss the advantages and disadvantages of such data. In addition, a dataset has been created for the experimental needs.
- Creation of novel techniques allowing polarimetric data to be augmented by preserving the physical properties from this modality.
- Detection and segmentation of reflective areas through standard convolutional deep learning techniques.

The various past works on which this paper is based are presented in Section 2. Then, the different processes of our implementation are introduced in the Section 3. The forth section summarize all the necessary steps for the experiment. Also, this section presents the results of the two modalities used (polarimetry and RGB) and the discussions that will compare the results obtained and also their interpretations. The last section concludes on this work as well as offers an opening on future work.

2 RELATED WORKS

2.1 Scene Segmentation

The pixel-wise semantic segmentation is the ability of giving a label for each pixel of an image. This task requires an accurate learning of the features on a set of image. This leads to the creation of a generic model which is able to classify at the pixel-level. Many research proved that deep learning models tend to make complex task learning and understanding accessible. Computer vision has benefited from the advances in this field to progress in general tasks. More precisely, many applications of semantic segmentation has been developed; among the most represented: road scene segmentation (Oliveira et al., 2016), indoor scene understanding (Gupta et al., 2014; Qi et al., 2017).

The first remarkable deep learning based segmentation is the FCN from Long et al. (Long et al., 2015), that allows the segmentation of image of any sizes without fully connected layers. Starting from this previous paper, as the years and the evolution of power increased, multiple networks, each with better performance, have been released: SegNet (Badrinarayanan et al., 2015), DeepLab (Chen et al., 2015; Chen et al., 2016; Chen et al., 2018), Image-to-Image (Isola et al., 2017) , Conditional Generative Adversarial Networks (Wang et al., 2018).

2.2 Polarimetry

Polarimetry is the science of measuring the polarized state of the light. As a consequence, a polarimetric camera (Wolff and Andreou, 1995) gives the experience of recovering the light changes in the captured environment. Because of this behavior, the information from this camera could be the perfect candidate as a discriminant factor for complex scene semantic understanding.
As shown in Figure 2, polarimetric images can be used advantageously, because the reflection operates a direct impact on the image.
For example, Kai Berger et al. proposed a method for depth recovering from polarimetric images in urban environment (Berger et al., 2017), treating the modality as a common RGB camera. Other polarization based systems have been proposed for water detection using polarized information. For example, Nguyen et

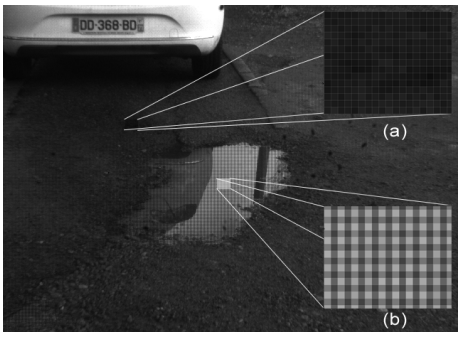

Figure 2: Reflection Influence on Polarimetry. (a) is a zoom on the non-polarized area and (b) on a polarized area. Clearly, on a polarized surface, the micro-grid appears and reveals an intensity change according to the polarizer affected.

al. proposed a method for water tracking with a polarized stereo system (Nguyen et al., 2017) achieving an approximate accuracy of 65% exceeding the previous state of the art method accuracy of approximately 45% (Yan, 2014). Rankin and Matthies proposed an application in recognition of mud for autonomous robotics and offered a full benchmark for the segmentation processes (Rankin and Matthies, 2010b).
One of the disadvantages of these previous methods is the lack of automation of tasks or the difficulty of deployment. In contrast, a deep learning approach allows the creation of a model that can be reused and redesigned as it goes along.

Despite the useful and informative aspects of polarimetric system, the use of such cameras have been quite restricted, due to the limitation of hardware and automatic integration. Using the DoFP technique (Nordin et al., 1999b; Nordin et al., 1999a; Millerd et al., 2006), the polarimetric camera has been introduced, which allow easier integration. DoFP technique allows having the polarized filters in an array directly on the sensor. In this design, four polarized filters, with unique angles, are used to capture four different measurements instantly in one shot. Many image processing and computer vision applications can benefit from recent DoFP-polarimetric camera.

In this paper, we are introducing polarimetry to the field of pixel-wise semantic segmentation for outdoor scenes.

3 METHOD

3.1 Polarimetric data pre-processing

Contrary to other standard type of images (RGB, gray-scale, etc.), the image provided by a DoFP camera is composed of 2x2 super-pixels. Consequently, we use an interpolation method (Ratliff et al., 2009) in order to recover polarimetry images. The key idea behind this transformation is to extract three one-channel images to represent three physical notions: the Angle of Polarization (AoP), the Degree of Polarization (DoP) and the Intensity (I). The AoP represents the value of the angle of polarization at each pixel while the DoP is the strength of the polarization state of the incoming light for each pixel.
In nature, the light is mainly partially linearly polarized which reduces the Stokes parameters to three parameters as bellow:

$$S = \begin{pmatrix} s_0 \\ s_1 \\ s_2 \\ 0 \end{pmatrix} = \begin{pmatrix} P_0 + P_{90} \\ P_0 - P_{90} \\ P_{45} - P_{135} \\ 0 \end{pmatrix}, \tag{1}$$

where $s_{\{0,1,2\}}$ are the three-first Stokes parameters, and $P_{\{0,45,90,135\}}$ the intensity output images corresponding to the orientation of the polarizer. The commonly used Stokes vectors can be normalized by s_0:

$$\bar{S} = \begin{pmatrix} \bar{s_1} \\ \bar{s_2} \\ 0 \end{pmatrix} = \frac{1}{s_0} \begin{pmatrix} s_1 \\ s_2 \\ 0 \end{pmatrix}. \tag{2}$$

AoP and DoP can be deduced according to:

$$\text{DoP} = \sqrt{\bar{s_1}^2 + \bar{s_2}^2}, \tag{3}$$

$$\text{AoP} = \frac{1}{2}\tan^{-1}(\frac{s_1}{s_2}). \tag{4}$$

The last parameter I is the intensity which is the combination of all polarized states intensities:

$$\text{I} = \frac{P_0 + P_{45} + P_{90} + P_{135}}{2}. \tag{5}$$

After this computation, three gray-scale description images of the raw polarimetric data are obtained. We have chosen to build an HSL (Hue Saturation Luminance) image mapping the three previous sources of information. This colorspace allows specific behavior per channel which fit with the data provided by AoP, DoP and I. The hue is commonly a 360° periodic value, the saturation is a value between zero and one as well as the value for the luminance. To fit the prerequisites of this color space, we made the adaptation and/or normalization of our images according to each channel and then merged them together (Wolff and Andreou, 1995).

$$H \longrightarrow 2 * \text{AoP}, \quad S \longrightarrow \text{DoP}, \quad L \longrightarrow I/255. \tag{6}$$

HSL can be seen as a single 3-channel image. This allows any RGB pre-initialized DL network to deal with these images. It is then possible to augment the data taking advantage of the HSL representation.

3.2 Polarimetric data augmentation

As previously explained, polarimetric information characterizes the vectorial representation of light. By consequence, any image has a unique meaning only for these precise camera parameters and orientations. The augmentation procedure consists in creating new images with the application of a transformation and/or an interpolation. The constraints induced by the type of data are exported to any transformation applied. The luminance and saturation channels can be released of the constraints because their attributed values are invariant around the optical axis. Contrarily, the hue is affected by this transformation. It is necessary to recompute the hue coherently with the physical properties of the camera. In this unique case, the angle of polarization will have a consistent physical meaning.

While rotating the camera counter-clockwise, the angle of polarization is rotated clockwise. Let θ be the applied rotation angle to the camera, R_θ the rotation matrix and H the hue channel of the image:

$$\mathrm{H}_{\text{rotated}} = R_\theta(\mathrm{H}_{prev} - \theta). \tag{7}$$

At the end of this computation, the image will keep its physical properties and be rotated.
As shown in Table 1, a set of transformations has been developed to give the ability to extend any polarimetric images dataset and it is remarkable that only the hue channel needs some modifications to stick to physical properties. The translation is only a shift in the images, which means that there is no modification in the view point of the camera. Since a polarimetric camera is dependent on the actual position and view point, the hue channel remains invariant to translation. On the other hand, if the camera lens has a wide angle, then in this case an additional transformation will be necessary (Table 1 -*).

3.3 Pixel-wise Segmentation with Deep Learning

Deep learning shows great performances on learning new kind of features and giving genericity to a model.

SegNet (Badrinarayanan et al., 2015) is employed in our work because of its robustness and short training time. The SegNet has an encoder-decoder design and an architecture composed of 36 layers. In our application, the key point in this design lies in the encoder part. It is composed of 13 layers, fitting perfectly the VGG-16 (Simonyan and Zisserman, 2014) ConvNet configuration B. In consequence, a transfer learning (Pan et al., 2010; Torrey and Shavlik, 2010) method can be applied allowing pre-initialization of the network. Considering this approach, an efficient training can be operated, avoiding a costly end-to-end training.

3.4 A New Dataset: PolaBot

Acquisition was conducted to provide a new multi-modal dataset PolaBot with polarimetric images. To the best of our knowledge, no such specific dataset has been released yet. Moreover, in order to make this dataset reliable for different fields (robotic, autonomous navigation, etc.), the acquisitions were made with a multi-modal system of four calibrated cameras. Three synchronized modalities are represented, two RGB from different angles, one NIR (Near-Infrared) and one polarimetric camera. In addition, this collection of information will allow a strong and efficient benchmark, giving the opportunity to compare standard modality to the polarimetry for the exact same scenes and application. This dataset is available at: http://vibot.cnrs.fr/polabot.html.

4 EXPERIMENTS

To confirm our hypothesis of the polarimetric data being more efficient than standard modality for our application, experiments have been conducted, allowing a comparison.
All the experiments were performed on the same dedicated server composed of an Nvidia Titan Xp (12GB Memory) GPU, 128GB of RAM and two CPU accumulating a total of 24 physical cores (48 threads).

For the SegNet Network, internal parameters of the training must be set. We had to set the loss function and the optimizer. We decided to use Adam (Kingma and Ba, 2014) as optimizer and as the loss function the cross entropy loss, defined as:

$$CEL(p,q) = -\sum_{\forall x} p(x) log(q(x)), \tag{8}$$

where x represents the class, $p(x)$ is the prediction for the x class and $q(x)$ the ground truth. Also, for all the training, the learning rate was initialized as 10^{-4} and a maximum of 500 epochs.

Table 1: Augmentation procedure per channels. Here "-" represents invariant, "*" represents that under condition this parameter can be modified.

	AoP (H)	**DoP (S)**	**Intensity (L)**
Crop	-	-	-
Roation	$R_\theta(\mathrm{H}-\theta) \pmod{360}$	-	-
Symmetry (Flip)	$-H \pmod{360}$	-	-
Translation	-*	-*	-

4.1 Metrics

To measure the efficiency of the training, common metric has been employed during the process: MIoU (Mean Intersection over Union), F1 Score, Mean Accuracy and Overall Accuracy. The IoU is defined as:

$$\mathrm{IoU} = \frac{\text{Area of Intersection}}{\text{Area of Union}}. \tag{9}$$

Another widely used metric is the F1 score. This metric observes the same behavior as the MIoU since the perfect score is 1. This metric is a combination of the recall and the precision, which correspond respectively to the relevance and the robustness of the results:

$$\text{F1 Score} = 2 * \frac{\text{precision} \,.\, \text{recall}}{\text{precision} + \text{recall}}. \tag{10}$$

Finally, the per-class accuracy is the measurement of fitting for each class:

$$\text{Accuracy}_C = \frac{\sum_i [\, p(i) = C \cap GT(i) = C \,]}{\sum_i [\, GT(i) = C \,]}, \tag{11}$$

where C is the class, $p(i)$ is the predicted class of pixel i and $GT(i)$ the ground truth.

4.2 Results

A color chart is used, therefore, for the next images, each area color in the image will have a meaning shown in the Table 2.

Each class has a clear meaning except unlabeled and None. None corresponds to zones segmented by hand but considered non-revealing with respect to our application. The unlabeled class, on the other hand, comes from manual segmentation errors. This class is the eighth class but is not necessarily consistent. Therefore, the results for this class will be neglected and taken into account in the conclusions drawn.

4.2.1 Training Results

Metrics for each epoch has been computed. This procedure allows seeing the fitness evolution of the model.

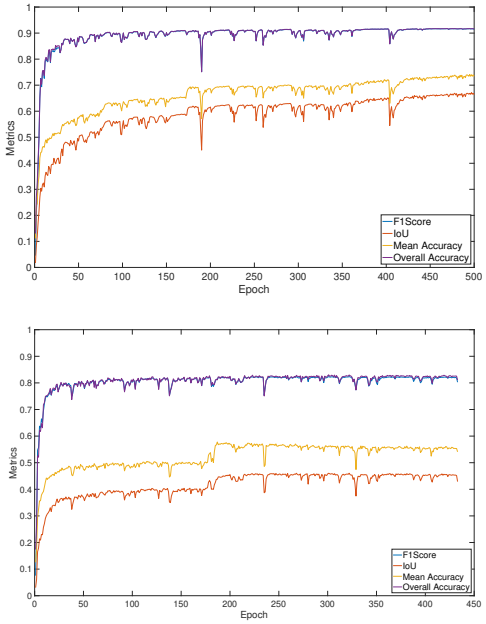

Figure 3: Training Results - Top is the graph corresponding to metrics estimation for the polarimetric data while training. The bottom graph corresponds to the RGB data training.

As shown in Figure 3, both curves are different according to the data provided to the network. First, it is possible to see that the two processes did not stop at the same time. While the network with polarimetric data reached 500 epochs, the network processing RGB data ended at 432 epochs. Indeed, we had put into place a stopping criterion to prevent the network from decaying. This result means that the SegNet RGB has experienced a decrease in its validation metrics for more than 10 epochs. However, our process allows the recovery of the optimal state in order to assess the so-called "optimal" results.

In a second step, it is possible to notice the differences in metric values. The SegNet Polarimetry reaches a MIoU value of 0.66, an F1 score of 0.91 and an average accuracy of 0.73. On the other hand, the SegNet RGB appears to be less efficient with lower scores:

Table 2: Color chart. This color chart allows uniformity in the visualization of results (each class has an affiliated color).

Meaning	Unlabeled	Sky	Water	Windows	Road	Cars	Building	None
Color	Black	Green	Blue	Yellow	Orange	Red	Grey	White

MIoU of 0.42, F1 score of 0.8 and average accuracy of 0.54.
It is possible to conclude this estimate of training by stating that SegNet Polarimetry seems to perform better during the learning phase.

4.2.2 Testing Results

The testing results correspond to the results obtained at the output of the network. As shown in the Table 3, in order to compare the impacts of each type of data, their respective accuracy by class was calculated for RGB and polarimetry and followed by comparison via difference:

$$\text{Accuracy}_{\text{Diff}} = \text{Accuracy}_{\text{Pol}} - \text{Accuracy}_{\text{RGB}}. \quad (12)$$

The Figure 4 shows the results obtained at the output of the SegNet Polarimetry and the Figure 5 those of the SegNet RGB. The segmentation is correct in both cases and visually offers good results.

4.3 Discussion

As shown in the Table 3, very high accuracy can be observed in all segmented classes using polarimetric data. As the data set is not generic, the sky remains on the same tone (blue), which gives a significant advantage over the RGB mode. The other classes where the RGB model is better are: road, water and none. These differences are minimal and can be explained in several ways. One of our hypotheses concerns the difference in manual segmentation for ground truth. RGB and polarimetry were segmented independently, increasing uncertainties. The difficulty of segmentation of certain classes must be taken into account. Another way to look at these results is to consider the advantages and disadvantages of cameras in relation to the dataset. For example, the road can be polarized if there is a high temperature; therefore, polarimetry would have an advantage over the RGB model. Since the dataset is acquired in only one type of weather condition, the RGB may have an advance over the other model, which may explain these results.
However, polarimetry model gives very high accuracy in all the classes. More precisely, when segmenting areas such as windows, cars and building, the model obtain a big positive difference compared to the RGB. The window segmentation is almost twice more performant using polarimetry model than RGB model. Indeed, these results can be explained by the polarization state of such areas.

5 CONCLUSION AND FUTURE WORK

In this paper, we proposed the introduction of polarimetry to pixel-wise road scenes segmentation field. Since to our knowledge there was no dataset with outdoor scenes captured via polarimetry, we created our own dataset. This dataset being made up of several modalities, the key idea was to have a comparison measure. As polarimetric data require meticulous exploitation, we have developed an augmentation method to preserve the physical properties of this modality. This approach defines the possible transformations and provides the necessary formulas for a rotation or flipping. We then used our augmented dataset as input to the SegNet Network to estimate the results. After comparing the SegNet Polarimetry and the SegNet RGB we can deduce that polarimetry offers a considerable advantage over RGB. Indeed, reflective areas are better detected while maintaining or improving the segmentation performance of other areas. We can conclude that polarimetry can provide a new type of information usefull in many fields such as robotics, computer vision or autonomous cars.

However, there are still some areas for improvement. One area for improvement is the use of a more complex network with deeper and more abstract functionalities. This will then allow the results to be compared between a simple network and a deeper network. The immediate objective of improvement is to use raw polarimetric images to eliminate any preprocessing.

ACKNOWLEDGEMENTS

This work was supported by ANR VIPeR, ANR ICUB. We gratefully acknowledge the support of NVIDIA Corporation with the donation of GPUs used for this research.

Table 3: Per-class Accuracy and accuracy Difference.

	Sky	Water	Windows	Road	Cars	Building	None	Mean
Polarimetry	75.34 %	75.70 %	82.85 %	77.82 %	71.40 %	87.69 %	78.95 %	78.54 %
RGB	89.57 %	78.61 %	44.50 %	78.45 %	48.48 %	67.84 %	83.4 %	69.83 %
Difference	-14.23%	-3.51 %	38.35 %	-0.63 %	22.92 %	19.85 %	-4.45 %	8.71 %

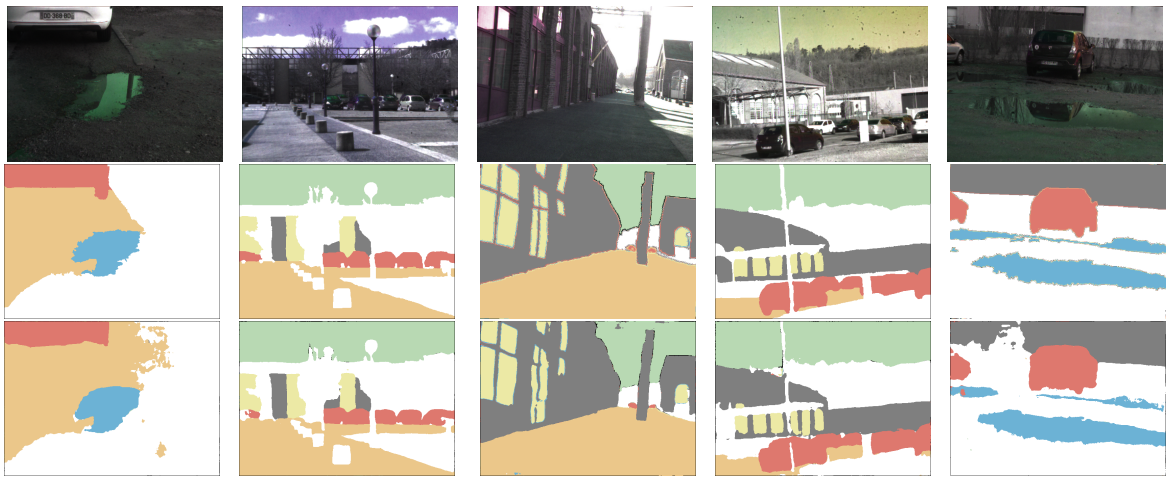

Figure 4: Polarimetry Results - Test Set Output.The top row is the input HSL image. The middle row is the ground truth manually segmented. The bottom row is the prediction output by the SegNet Polarimetry.

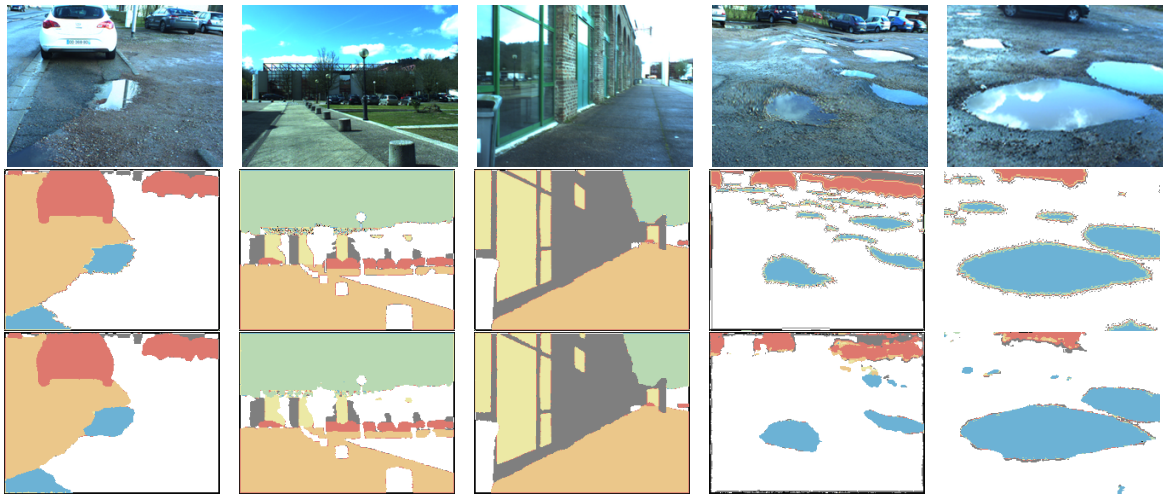

Figure 5: RGB Results - Test Set Output. The top row is the input RGB image. The middle row is the ground truth manually segmented. The bottom row is the prediction output by the SegNet RGB.

REFERENCES

Badrinarayanan, V., Handa, A., and Cipolla, R. (2015). Segnet: A deep convolutional encoder-decoder architecture for robust semantic pixel-wise labelling. *CoRR*, abs/1505.07293.

Berger, K., Voorhies, R., and Matthies, L. H. (2017). Depth from stereo polarization in specular scenes for urban robotics. In *Robotics and Automation (ICRA), 2017 IEEE International Conference on*, pages 1966–1973. IEEE.

Chen, L.-C., Papandreou, G., Kokkinos, I., Murphy, K., and Yuille, A. L. (2015). Semantic image segmentation with deep convolutional nets and fully connected crfs. In *ICLR*.

Chen, L.-C., Papandreou, G., Kokkinos, I., Murphy, K., and Yuille, A. L. (2016). Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *arXiv:1606.00915*.

Chen, L.-C., Papandreou, G., Kokkinos, I., Murphy, K., and Yuille, A. L. (2018). Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE transactions on pattern analysis and machine intelligence*, 40(4):834–

848.

Couprie, C., Farabet, C., Najman, L., and LeCun, Y. (2013). Indoor semantic segmentation using depth information. *arXiv preprint arXiv:1301.3572*.

Gupta, S., Girshick, R., Arbeláez, P., and Malik, J. (2014). Learning rich features from rgb-d images for object detection and segmentation. In *European Conference on Computer Vision*, pages 345–360. Springer.

Isola, P., Zhu, J.-Y., Zhou, T., and Efros, A. A. (2017). Image-to-image translation with conditional adversarial networks. *arXiv preprint*.

Kingma, D. P. and Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

Long, J., Shelhamer, E., and Darrell, T. (2015). Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3431–3440.

Millerd, J., Brock, N., Hayes, J., North-Morris, M., Kimbrough, B., and Wyant, J. (2006). Pixelated phase-mask dynamic interferometers. In *Fringe 2005*, pages 640–647. Springer.

Morel, O., Meriaudeau, F., Stolz, C., and Gorria, P. (2005). Polarization imaging applied to 3d reconstruction of specular metallic surfaces. In *Machine Vision Applications in Industrial Inspection XIII*, volume 5679, pages 178–187. International Society for Optics and Photonics.

Nguyen, C. V., Milford, M., and Mahony, R. (2017). 3d tracking of water hazards with polarized stereo cameras. In *Robotics and Automation (ICRA), 2017 IEEE International Conference on*, pages 5251–5257. IEEE.

Nordin, G. P., Meier, J. T., Deguzman, P. C., and Jones, M. W. (1999a). Diffractive optical element for stokes vector measurement with a focal plane array. In *Polarization: Measurement, Analysis, and Remote Sensing II*, volume 3754, pages 169–178. International Society for Optics and Photonics.

Nordin, G. P., Meier, J. T., Deguzman, P. C., and Jones, M. W. (1999b). Micropolarizer array for infrared imaging polarimetry. *JOSA A*, 16(5):1168–1174.

Oliveira, G. L., Burgard, W., and Brox, T. (2016). Efficient deep models for monocular road segmentation. In *Intelligent Robots and Systems (IROS), 2016 IEEE/RSJ International Conference on*, pages 4885–4891. IEEE.

Pan, S. J., Yang, Q., et al. (2010). A survey on transfer learning. *IEEE Transactions on knowledge and data engineering*, 22(10):1345–1359.

Qi, C. R., Su, H., Mo, K., and Guibas, L. J. (2017). Pointnet: Deep learning on point sets for 3d classification and segmentation. *Proc. Computer Vision and Pattern Recognition (CVPR), IEEE*, 1(2):4.

Rahmann, S. and Canterakis, N. (2001). Reconstruction of specular surfaces using polarization imaging. In *null*, page 149. IEEE.

Rankin, A. and Matthies, L. (2010a). Daytime water detection based on color variation. In *Intelligent Robots and Systems (IROS), 2010 IEEE/RSJ International Conference on*, pages 215–221. IEEE.

Rankin, A. L. and Matthies, L. H. (2010b). Passive sensor evaluation for unmanned ground vehicle mud detection. *Journal of Field Robotics*, 27(4):473–490.

Ratliff, B. M., LaCasse, C. F., and Tyo, J. S. (2009). Interpolation strategies for reducing ifov artifacts in microgrid polarimeter imagery. *Optics express*, 17(11):9112–9125.

Simonyan, K. and Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.

Torrey, L. and Shavlik, J. (2010). Transfer learning. In *Handbook of Research on Machine Learning Applications and Trends: Algorithms, Methods, and Techniques*, pages 242–264. IGI Global.

Wang, T.-C., Liu, M.-Y., Zhu, J.-Y., Tao, A., Kautz, J., and Catanzaro, B. (2018). High-resolution image synthesis and semantic manipulation with conditional gans. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, volume 1, page 5.

Wolff, L. B. and Andreou, A. G. (1995). Polarization camera sensors. *Image and Vision Computing*, 13(6):497–510.

Yan, S. H. (2014). Water body detection using two camera polarized stereo vision. *International Journal of Research in Computer Engineering & Electronics*, 3(3).