

# POLARIMETRIC IMAGE AUGMENTATION

Marc Blanchon<sup>1</sup>, Olivier Morel<sup>1</sup>, Fabrice Meriaudeau<sup>1</sup>, Ralph Seulin<sup>1</sup> and Désiré Sidibé<sup>2</sup>

<sup>1</sup> ERL VIBOT CNRS 6000, ImViA  
Université Bourgogne Franche Comté  
Le Creusot, France

<sup>2</sup> Université Paris-Saclay  
Univ Evry, IBISC  
91020, Evry, France

## ABSTRACT

Robotics applications in urban environments are subject to obstacles that exhibit specular reflections hampering autonomous navigation. On the other hand, these reflections are highly polarized and this extra information can successfully be used to segment the specular areas. In nature, polarized light is obtained by reflection or scattering. Deep Convolutional Neural Networks (DCNNs) have shown excellent segmentation results, but require a significant amount of data to achieve best performances. The lack of data is usually overcome by using augmentation methods. However, unlike RGB images, polarization images are not only scalar (intensity) images and standard augmentation techniques cannot be applied straightforwardly. We propose to enhance deep learning models through a regularized augmentation procedure applied to polarimetric data in order to characterize scenes more effectively under challenging conditions. We subsequently observe an average of 18.1% improvement in IoU between non augmented and regularized training procedures on real world data.

*Index Terms*— Polarimetry, Augmentation, Deep Learning, Computer Vision

## 1. INTRODUCTION

Navigation in urban environments can be prone to errors due to highly reflective areas while using RGB cameras. On the other hand, polarization imaging can cope with such environments. Data driven segmentation methods using polarimetric images as input require a large and representative data-set illustrating hazardous areas such as cars, water spillage, windows, etc. Unfortunately, very few polarimetric datasets are available or they do not consider scenes from urban areas. A common approach to overcome the lack of sufficient data in training deep CNNs is to use augmentation methods to increase the amount of data and make learning more generic and therefore avoid overfitting [1]. However, unlike RGB images, polarization images are not only scalar (intensity) images and standard augmentation techniques cannot be applied straightforwardly.

In this paper, we consider the problem of artificially augmenting polarization images that are in direct relation to the physics of the scene acquired by the camera. This acquisition/scene relationship drastically complicates the expansion of a dataset with standard techniques. However, it is necessary to address this issue, as the available data on polarimetric modality is extremely limited despite its increasingly popular uses [2–5].

We have explored augmentation operations applicable to polarimetry under any condition. Initially, using the interpolation proposed by Ratliff et al. [6] combined with the channel organization proposed by Wolff and Andreou [7], we propose an approach for

applying rotation and/or symmetry to polarimetric images. In our case, for the experimental results, multiple trainings were performed with either a raw dataset (limited number of images), a standardly increased dataset (significant number of images without respecting physical properties) or an augmented dataset following our procedure (significant number of images with unaltered physical properties). To focus exclusively on the impact of the data, all trainings were performed using the DeepLabV3+ network [8]. Furthermore, we demonstrate the integrity of physical properties and their effectiveness for segmentation purposes.

The remainder of this paper is organized as follows. Section 2 provides a brief overview of related works including pixel-wise segmentation and polarimetric image analysis. Then, in Section 3 the proposed augmentation method is described in details. Section 4 presents experimental results showing the effectiveness of our approach, and the paper ends with concluding remarks in Section 5.

## 2. RELATED WORKS

### 2.1. Pixel-Wise Semantic Segmentation

Most of the research on pixel-wise semantic segmentation use conventional images as input data, either RGB or depth images. We can observe a constant evolution of networks accuracy due on the one side to the improvement of networks and on the other side to the increase in datasets sizes. Also, some tasks are recurrent in the community: segmentation of urban scenes [1, 8–13], indoor scenes understanding [14–17] or medical images analysis [18–20]. The task addressed in this paper shares a common aspect with medical imaging. Indeed, one of the common disadvantages of medical images and urban scene acquired with polarimetric sensor is the lack of large annotated datasets for training. In particular, the specificity of the polarimetric information makes it rich but also rare. Semantic segmentation being a dense and valuable information, capable of characterizing a scene at any point, it can be exploited very extensively in robotics for navigation and detection.

### 2.2. Polarimetric Modality

Polarimetry has unique properties since it allows the acquisition of changes in the state of light [7]. It particularly characterizes reflective areas. The attractiveness of these images is their ability to describe both the diffuse and specular parts of a scene. Considering all these advantages, polarization imaging offers a comprehensive range of possibilities that could enhance more standard modality such as RGB.

As polarization images not exclusively carry intensity information but also information about the surface it is reflected upon, these extra information make it possible to extend the range of applications



**Fig. 1:** Rotation on real image and impact. From left to right, the initial image, then the  $+90^\circ$  rotated image and then the regularized image using the equations 1 and 3.

in computer vision, like Shape from Polarimetry, which reconstructs specular (or partially specular) objects [3, 21–23]. Other works have been carried out to improve certain approaches, such as attitude estimation [4, 24], water hazard detection [5, 25], catadioptric camera calibration [26], and depth estimation [2, 27].

Considering applications in urban environments, inferring extra knowledge about objects from the reflected light, would result in better perception and therefore useful for unmanned vehicles and robots for example.

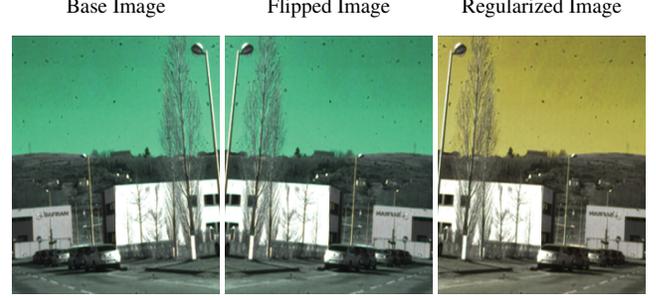
### 3. METHOD

#### 3.1. Reminder on polarimetric image processing

Polarimetric images are obtained either by rotating a polarizer in front of the camera or via a sensor exploiting the division of focal plane (DoFP). In both cases, pre-processing is frequently necessary to acquire the informative part of the images. Indeed, raw images provide very low level description of the scenes. Commonly, many approaches require a transformation either to have specific information such as the angle of polarization (AoP) and the degree of polarization (DoP), or to interpolate the images. Ratliff et al. [6] proposed a widely used approach to merge these two information. Four images acquired with four distinct polarizer angles ( $0^\circ$ ,  $45^\circ$ ,  $90^\circ$ ,  $135^\circ$ ) are required for this pre-processing step, via four unique acquisitions or using a DoFP sensor. From these 4 images,  $P_0$ ,  $P_{45}$ ,  $P_{90}$ ,  $P_{135}$ , one can reconstruct three informative images: the intensity  $I$ , the angle of polarization AoP and the degree of polarization DoP. The last two images correspond, from a physical point of view, to the angle and power of reflection, respectively [28]. These informative images are combined as proposed by Wolff and Andreou [7] for display into an HSL image. As shown in Fig.1, first image, the singular characteristic of these new images is that the more reflective the surface is the more colored it appears.

#### 3.2. Applying transformations to polarimetric images

Once the images have been transformed it is possible to apply augmentation procedures. However, care must be taken to consider the physical properties induced by the scene (and also by the sensor). Here, we use a camera with a division of focal plane, which makes it easier to explain the influence of rotation on images. Indeed, the pixel organization due to the sensor allows the preservation of the integrity of the transformations by validating their order. To however apply a rotation to the image, we need an additional process to "re-align" the pixels and keep the physical properties of the polarization angle.



**Fig. 2:** Flipping on real image and impact. From left to right, initial image, the flipped image and the regularized image using the equations 2 and 3.

In this case, we apply an anti-clockwise rotation to the image while at the same time applying a clockwise rotation to the polarization angle, that brings back the physics of the scene. Let  $\theta$  be the rotation angle applied to the camera,  $R_\theta$  the rotation operation and  $H$  the hue channel of the image (which as a reminder corresponds to the AoP):

$$H_{\text{rotated}} = R_\theta(H_{\text{prev}} - 2 * 1\theta). \quad (1)$$

Starting from the sensor and more particularly from a set of 4 DoFP pixels, if a pure rotation is applied, then the sensor will be disoriented and therefore the physical direction of each image will not be respected. The regularization applied by the equation (1) then allows the image to be redirected and thus to keep the correct alignment of each polarizer. The other channels are independent of the physical properties, which allows to neglect them in this transformation step. These two channels  $S$  and  $L$  are still subjected to the rotation operation.

Another possibility of augmentation proposed in this paper is symmetry. To ensure the integrity of the physical properties of the image, symmetry is applied as follows:

$$H_{\text{flipped}} = -H_{\text{prev}}. \quad (2)$$

As before for rotation, we can graphically show in Fig.2 the impact of pure flipping and the action of regularization. Knowing that the Hue channel of images is 360 degrees periodic, flipping consists in reversing the selected axis for the transformation. The inversion of the values of the Hue channel is made possible by the use of the periodic 360 degree property by using consecutively the equations 2 and 3. This procedure is illustrated in Fig.2.

In both augmentation operations previously presented, it is necessary to maintain a consistency with respect to the properties of the color space. The Hue channel being a value between 0 and 360, it is then necessary to normalize the intensities:

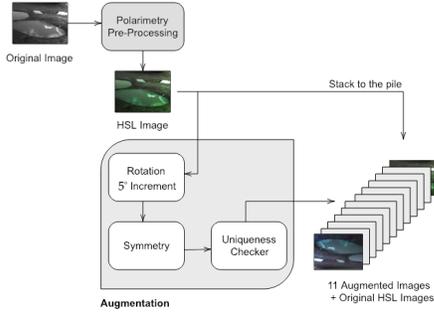
$$H_{\text{final}} = H_{\text{transformed}} \pmod{360} \quad (3)$$

where  $H_{\text{final}}$  is the final image and  $H_{\text{transformed}}$  is the image resulting from a rotation, symmetry or a combination of both.

## 4. EXPERIMENTS

#### 4.1. Implementation details

In this section, different results from several augmentation procedures on an unique video sequence are presented and compared.



**Fig. 3:** Illustration of the augmentation procedure per image. This process is repeated for each image in the original dataset to obtain a consistent large dataset. Then, the entire set of augmented images is shuffled.

The idea here is to prove the augmentation reliability and show its impact in improving segmentation results.

**Datasets:** The dataset is unevenly divided into a part dedicated to training and a part dedicated to validation. The validation set is composed of 50 images selected for their characteristics allowing to have a balanced representation of the classes in the images and thus to have more precise validation metrics. As for the rest of the dataset, it is employed for training and will not be balanced so that it is closer to real world data.

Initially, the training data set is composed of 178 images that are either kept as they are, or augmented in a standard way, without regularization, or augmented using our approach. Whether the dataset is increased with our method or not, the transformations involved are identical and the set accumulates a total of 2136 images. As shown in Fig.3, the augmentation is performed by applying a random rotation in increments of  $5^\circ$  and/or symmetry with a probability of 20%. Note that each augmentation is unique.

The test dataset is a video sequence comprised of 8,049 images acquired at a frequency of 10Hz sharing many characteristics with the training dataset. It was acquired by mounting a Trioptics Polar-Cam 4D Technology V polarimetric camera on a remotely operated Robotnik Summit XL robot and using the ROS operating system.

**Network training:** The experiments are conducted using a 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). We use DeepLabV3+ [8] network, either with pre-trained parameters or without. Indeed, it is equally significant to compare the influence of pre-training. Since we use DeepLabV3+, it is possible to pre-train the xception subnetwork using the most recent provided model<sup>1</sup>. The hyper-parameters and the loss function are kept identical for both models (pre-trained or not). We set empirical parameters like epoch number to 150, learning rate to  $\eta = 10^{-2}$ , batch size to 8 and use Adam algorithm as optimizer.

With regard to the loss, here an adapted Sørensen-Dice index is used to take into account the fact that some classes are under-represented in the dataset:

$$\lambda = \frac{\sum_c^N 1 - \frac{2|X_c \cap Y_c|}{|X_c| + |Y_c|}}{N}, \quad (4)$$

<sup>1</sup><https://data.lip6.fr/cadene/pretrainedmodels/>

with  $X$  the label,  $Y$  the prediction,  $c$  the class and  $N$  the number of classes. Since classes are unequally represented in the dataset, this metric allows an equal valuation of each of them unlike other losses.

**Metrics:** Apart from qualitative evaluation such as Fig. 4, we have calculated a comprehensive range of metrics to quantitatively compare the different models according to augmentation processes. Results are shown in Tab. 1 and correspond to metrics computed on a sub-set composed of 15 different images where all classes are represented. This avoids redundancies and therefore allows for a more detailed case-by-case examination. The scenes correspond to urban areas with seven different classes: *Road* (dark yellow/orange), *Buildings* (grey), *Cars* (red), *Water* (blue), *Windows* (light yellow), *Sky* (green) and *None* (light grey).

## 4.2. Discussion

In this section, we will discuss the different impacts of the augmentation on the network. As shown in Fig. 4, a brief visual assessment makes it difficult to observe substantial differences between the models even if the regularized procedure seems more appropriate. It is noteworthy that, when using the proposed augmentation method, visual aberrations are reduced. A predominant visual defect of all models is the absence of building detection. This will be discussed further below.

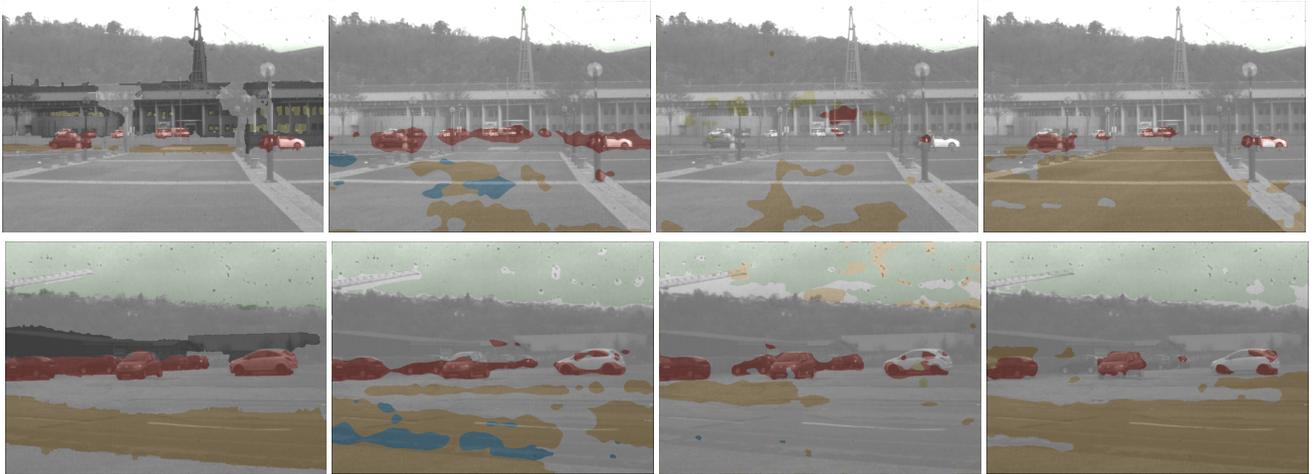
As shown in Tab. 1, a large panel of metrics has been calculated to evaluate the performance of each model. The two major metrics are mean intersection over union and recall.

For each of the major metrics, we have selected three classes representing reflective areas for comparison: *Water*, *Windows* and *Cars*. Class-specific metrics are, from a robotics application perspective, the critical points and also represent the core objective since they are derived from the detection of danger zones. In addition, two more general metrics are proposed to evaluate the models from various angles: precision and specificity.

From a general point of view, the Tab. 1 points out that the regularized augmentation allows for better results in the vast majority of cases. In more detail, the IoU shows that models with an appropriate dataset perform better and mainly in class-specific metrics where we can observe substantial differences. On the contrary, the standard augmentation produces the worst results in terms of IoU. Metrics also emphasize that it is better to use polarimetric data without augmentation than to augment them by neglecting physical properties of the scene. Indeed, with an unadapted augmentation method, the mean IoU is about twice lower than the result without augmentation. With the proposed augmentation approach, the mean IoU is improved by more than 40%.

The recall considers the model's ability to correctly classify zones independently of bad assignments. In our case, the recall ratio per class indicates the ability to perceive hazardous areas in general. Once more, the majority of the high results are obtained by the model with regularized data while the unsatisfactory results are held by the trained model with standard augmentation. However, for the class *Water*, we observe that the model without augmentation achieves a better recall. This can be explained by the tendency for this model to overuse this class (this phenomenon is visible Fig. 4, row 2 columns 1 and 4).

As stated above, the common defect of the six models compared is the inability to detect buildings. This difficulty can be explained by the "physical similarity" of the components forming the *None* and



**Fig. 4:** Examples of segmentation results according to the augmentation methods. From top to bottom are present the ground-truth, then consecutively the results from model with no pre-training with: no augmentation, standard augmentation and regularized augmentation.

**Table 1:** Impact of the augmentation procedure on DeepLabV3+ network. Specific classes have been highlighted in relation to the robotic application to witness the obstacle-wise performance. Due to the limited training, *Buildings* are almost undetected. For this reason, the averages denoted  $\setminus B$  exclude the *Buildings* class from the calculation.

Augmentation	PreTraining	IoU (%)					Recall (%)					Precision (%)	Specificity (%)
		@water	@windows	@cars	Mean	Mean $\setminus B$	@water	@windows	@cars	Mean	Mean $\setminus B$		
None	No	40.0	20.6	20.8	30.5	32.2	35.2	15.8	22.5	<b>50.9</b>	50.0	50.0	89.6
	Yes	54.0	10.3	43.46	33.5	34.8	<b>42.4</b>	15.3	57.4	43.3	50.3	50.1	91.0
Standard	No	0.1	3.4	12.4	14.8	13.1	35.0	25.8	15.0	31.8	28.0	41.7	88.7
	Yes	10.2	3.0	19.7	21.8	20.0	35.2	22.9	23.4	37.0	33.4	41.2	91.2
Regularized	No	63.9	13.3	46.7	<b>43.4</b>	<b>50.3</b>	39.2	21.9	<b>60.8</b>	43.4	<b>50.5</b>	48.5	<b>91.3</b>
	Yes	<b>70.0</b>	<b>26.6</b>	<b>47.1</b>	37.8	38.5	35.0	<b>26.0</b>	48.0	42.0	38.5	<b>53.7</b>	90.7

*Buildings* classes. As the training was limited, it is likely that more training epochs would have benefited for this specific class. In our approach to detection for robotics, this defect is uncritical since the class is not drastically classified as a danger. Plus, it can be deduced by exploiting the other correctly segmented classes allowing the implementation of a system of rules or constraints.

We can conclude the discussion by arguing that polarimetry allows for better detection of areas prone to reflection. This capability could benefit robotic applications by improving existing algorithms. Indeed, since reflections are characterized upstream of the network, the new learnt features are specific to the reflectivity of the surfaces in the scene. This hypothesis is validated by the unsatisfactory results obtained while neglecting physical properties of the modality and moreover when analyzing the segmentation of reflective areas like cars and waters.

## 5. CONCLUSION

Highly reflective areas induce errors while navigating using RGB sensors. To overcome this issue, one solution is to rely on polarimetric images; however, this solution also suffers from the lack of images to train deep models and we therefore develop a successful data augmentation technique that take into account the vector aspect of polarimetric images. We proposed several regularization processes

that maintain the integrity of the physical properties involved in polarization imaging. The experimental results show that the process of augmentation is useful and required when data amount is insufficient, but also that the results obtained at the end of the pipeline are improved in a significant way. To go further and increase the algorithm's capabilities, other transformations such as distortion and noise addition can be investigated. Nonetheless, the use of a new modality could be beneficial for robotics and autonomous vehicles. It has been demonstrated that, with polarimetric modality, important areas for these domains can be segmented more effectively if physical properties are unaltered. Moreover, it is also possible to reconsider the structure of the CNNs used to integrate the pre-processing step in order to avoid the HSL mapping of the polarized components.

## Acknowledgments

This work was supported by the French National Research Agency through ANR ICUB (ANR-17-CE22-0011). We gratefully acknowledge the support of NVIDIA Corporation with the donation of GPUs used for this research.

## 6. REFERENCES

- [1] Luis Perez and Jason Wang, "The effectiveness of data augmentation in image classification using deep learning," *CoRR*, vol. abs/1712.04621, 2017.
- [2] Kai Berger, Randolph Voorhies, and Larry H Matthies, "Depth from stereo polarization in specular scenes for urban robotics," in *Robotics and Automation (ICRA), 2017 IEEE International Conference on*. IEEE, 2017, pp. 1966–1973.
- [3] Zhaopeng Cui, Jinwei Gu, Boxin Shi, Ping Tan, and Jan Kautz, "Polarimetric multi-view stereo," in *Proc. of Computer Vision and Pattern Recognition (CVPR)*, 2017.
- [4] Mojdeh Rastgoo, Cédric Demonceaux, Ralph Seulin, and Olivier Morel, "Attitude estimation from polarimetric cameras," in *IEEE/RSJ Int. Conf. on Intelligent Robots and Systems-IROS*, 2018.
- [5] Chuong V Nguyen, Michael Milford, and Robert Mahony, "3d tracking of water hazards with polarized stereo cameras," in *Robotics and Automation (ICRA), 2017 IEEE International Conference on*. IEEE, 2017, pp. 5251–5257.
- [6] Bradley M Ratliff, Charles F LaCasse, and J Scott Tyo, "Interpolation strategies for reducing ifov artifacts in microgrid polarimeter imagery," *Optics express*, vol. 17, no. 11, pp. 9112–9125, 2009.
- [7] Lawrence B Wolff and Andreas G Andreou, "Polarization camera sensors," *Image and Vision Computing*, vol. 13, no. 6, pp. 497–510, 1995.
- [8] Liang-Chieh Chen, Yukun Zhu, George Papandreou, Florian Schroff, and Hartwig Adam, "Encoder-decoder with atrous separable convolution for semantic image segmentation," pp. 801–818, 2018.
- [9] Adam Paszke, Abhishek Chaurasia, Sangpil Kim, and Eugenio Culurciello, "Enet: A deep neural network architecture for real-time semantic segmentation," *arXiv preprint arXiv:1606.02147*, 2016.
- [10] Guosheng Lin, Anton Milan, Chunhua Shen, and Ian D Reid, "Refinenet: Multi-path refinement networks for high-resolution semantic segmentation.," in *Cvpr*, 2017, vol. 1, p. 5.
- [11] Liang-Chieh Chen, George Papandreou, Iasonas Kokkinos, Kevin Murphy, and Alan L Yuille, "Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs," *IEEE transactions on pattern analysis and machine intelligence*, vol. 40, no. 4, pp. 834–848, 2018.
- [12] Linhui Li, Bo Qian, Jing Lian, Weina Zheng, and Yafu Zhou, "Traffic scene segmentation based on rgb-d image and deep learning," *IEEE Transactions on Intelligent Transportation Systems*, vol. 19, no. 5, pp. 1664–1669, 2018.
- [13] Ziwei Liu, Xiaoxiao Li, Ping Luo, Chen Change Loy, and Xiaoou Tang, "Deep learning markov random field for semantic segmentation," *IEEE transactions on pattern analysis and machine intelligence*, vol. 40, no. 8, pp. 1814–1828, 2018.
- [14] Nathan Silberman, Derek Hoiem, Pushmeet Kohli, and Rob Fergus, "Indoor segmentation and support inference from rgb-d images," in *European Conference on Computer Vision*. Springer, 2012, pp. 746–760.
- [15] Mircea Cimpoi, Subhansu Maji, and Andrea Vedaldi, "Deep filter banks for texture recognition and segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 3828–3836.
- [16] Ankur Handa, Viorica Patraucean, Vijay Badrinarayanan, Simon Stent, and Roberto Cipolla, "Understanding real world indoor scenes with synthetic data," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2016.
- [17] Charles R Qi, Hao Su, Kaichun Mo, and Leonidas J Guibas, "Pointnet: Deep learning on point sets for 3d classification and segmentation," *Proc. Computer Vision and Pattern Recognition (CVPR)*, IEEE, vol. 1, no. 2, pp. 4, 2017.
- [18] Olaf Ronneberger, Philipp Fischer, and Thomas Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*. Springer, 2015, pp. 234–241.
- [19] Mohammad Havaei, Axel Davy, David Warde-Farley, Antoine Biard, Aaron Courville, Yoshua Bengio, Chris Pal, Pierre-Marc Jodoin, and Hugo Larochelle, "Brain tumor segmentation with deep neural networks," *Medical image analysis*, vol. 35, pp. 18–31, 2017.
- [20] Geert Litjens, Thijs Kooi, Babak Ehteshami Bejnordi, Arnaud Arindra Adiyoso Setio, Francesco Ciompi, Mohsen Ghafoorian, Jeroen Awm Van Der Laak, Bram Van Ginneken, and Clara I Sánchez, "A survey on deep learning in medical image analysis," *Medical image analysis*, vol. 42, pp. 60–88, 2017.
- [21] Stefan Rahmann and Nikos Canterakis, "Reconstruction of specular surfaces using polarization imaging," in *null*. IEEE, 2001, p. 149.
- [22] Olivier Morel, Christophe Stolz, Fabrice Meriaudeau, and Patrick Gorria, "Active lighting applied to three-dimensional reconstruction of specular metallic surfaces by polarization imaging," *Applied optics*, vol. 45, no. 17, pp. 4062–4068, 2006.
- [23] Olivier Morel, Fabrice Meriaudeau, Christophe Stolz, and Patrick Gorria, "Polarization imaging applied to 3d reconstruction of specular metallic surfaces," in *Machine Vision Applications in Industrial Inspection XIII*. International Society for Optics and Photonics, 2005, vol. 5679, pp. 178–187.
- [24] Abd El Rahman Shabayek, Cédric Demonceaux, Olivier Morel, and David Fofi, "Vision based uav attitude estimation: Progress and insights," *Journal of Intelligent & Robotic Systems*, vol. 65, no. 1-4, pp. 295–308, 2012.
- [25] Mohammad Iqbal, Morel Morel, and Fabrice Meriaudeau, "A survey on outdoor water hazard detection," *Skripsi Program Studi Siste Informatika*, 2009.
- [26] Olivier Morel, Ralph Seulin, and David Fofi, "Catadioptric camera calibration by polarization imaging," in *Iberian Conference on Pattern Recognition and Image Analysis*. Springer, 2007, pp. 396–403.
- [27] Achuta Kadambi, Vage Taamazyan, Boxin Shi, and Ramesh Raskar, "Polarized 3d: High-quality depth sensing with polarization cues," in *Proceedings of the IEEE International Conference on Computer Vision*, 2015, pp. 3370–3378.
- [28] Edward Collett, "Field guide to polarization," Spie Bellingham, WA, 2005.