Multi-Resolution and Multi-Domain Analysis of Off-Road Datasets for Autonomous Driving

Orighomisan Mayuku
Mechanical & Materials Engineering
and Ingenuity Labs Research Institute
Queen’s University
Kingston, ON Canada
o.mayuku@queensu.ca

Brian W. Surgenor
Mechanical & Materials Engineering
and Ingenuity Labs Research Institute
Queen’s University
Kingston, ON Canada
surgenor@queensu.ca

Joshua A. Marshall
Electrical & Computer Engineering
and Ingenuity Labs Research Institute
Queen’s University
Kingston, ON Canada
ORCID 0000-0002-7736-7981

Abstract—For use in off-road autonomous driving applications, we propose and study the use of multi-resolution local binary pattern texture descriptors to improve overall semantic segmentation performance and reduce class imbalance effects in off-road visual datasets. Our experiments, using a challenging publicly available off-road dataset as well as our own off-road dataset, show that texture features provide added flexibility towards reducing class imbalance effects, and that fusing color and texture features can improve segmentation performance. Finally, we demonstrate domain adaptation limitations in nominally similar off-road environments by cross-comparing the segmentation performance of convolutional neural networks trained on both datasets.

Index Terms—Features, domain adaptation, segmentation, vision for autonomous vehicles

I. INTRODUCTION

Contemporary approaches to autonomous vehicle navigation depend on significant amounts of labelled image data to train convolutional neural networks (CNNs) that are deployed in an end-to-end vision stack. Because CNNs learn by complex representations of shapes, they are well suited to urban environments that have distinct semantic classes, such as cars, person, street light etc. But, in off-road environments, classes are less distinct in shape because of the labelling strategy, which is influenced by the unique challenges encountered in off-road driving. One of these challenges is that, unlike in urban driving, the terrain cannot be taken for granted in off-road driving. Hence, terrain traversability is a key consideration [1]–[4], and in a previous work, we proposed and demonstrated a self-supervised end-to-end method for off-road navigation based on terrain traversal costs [5].

The traversability consideration means that, in off-road datasets, vegetation is often split into non-unique sub-classes, such as grass and traversable grass, based on perceived ease of traversability. Also, in urban driving there is a strict requirement to distinctly identify similar classes—for example, street lights from traffic lights—because the vehicle control strategy will be different in each case. Yet, this is not the case in remote off-road environments. Consequently, in off-road datasets, an obstacle class is often defined comprising an eclectic range of non-unique natural and man-made objects such as rocks of varying sizes and shapes, fences, power pylons, etc. A further complication is the natural class and spatial imbalance in the datasets; e.g., as shown in Fig. 1 for the Freiburg Forest dataset [6] in which obstacles are represented in more images than trees, but occupy fewer pixels showing spatial imbalance.

Whereas classification datasets such as ImageNet can contain millions of images, segmentation datasets contain fewer images because pixel-based labelling entails significant manual effort. What is more, off-road segmentation datasets contain even fewer images because remote off-road environments make data collection less convenient. For example, the Freiburg Forest dataset comprises only 366 images, compared to thousands of images in Cityscapes [7]. The limited size of off-road datasets, combined with the non-unique classes, makes semantic segmentation using CNNs with RGB-only inputs insufficient.

To specifically tackle the spatial imbalance problem, in

Fig. 1. Imbalance in the Freiburg Forest dataset [6]: the Tree and Obstacle classes are imbalanced; and obstacles appear in more images than trees, but have a smaller pixel count because they are physically smaller, showing spatial imbalance.
addition to augmentation techniques such as zooming, contemporary approaches use deeper networks with flexible convolutions [8], spatial pyramid pooling [9], and dilated convolutions [10] to increase the receptive field of the network in its deeper layers to ameliorate class imbalance effects. But the limited size of the off-road datasets reduces the benefits of these techniques.

One approach to improving segmentation performance is multi-modal fusion, whereby the RGB input is supplemented by other data sources such as infrared and depth data [6]. Apart from the fact that not all of these input modalities are intrinsic properties of the feature classes, e.g. depth, geometric matching of the different sensor inputs is a key issue. Similar pre-processing requirements, such as the need to de-noise stereo disparity maps, limit the scope for online application.

One option for fusion with RGB data that has not yet been explored in this end-to-end context, to the best of our knowledge, is texture. Although CNNs are typically believed to classify objects by shape, they have been shown to have a strong texture bias [11]. While texture has been used in classification tasks in remote sensing [12], [13], image matching [14] and supervised and unsupervised clustering on texture classification benchmarks [15], the texture descriptors have primarily been hand-crafted. Among texture descriptors, the local binary patterns (LBP) descriptor [16] is often preferred due to its ease of computation and robustness [17]. RGB-LBP fusion eliminates the need for geometric matching, which is necessary when using different sensor inputs; and LBP also enables multi-resolution analysis as in [18]. In this work, we study the effectiveness of RGB-LBP fusion.

Finally, off-road environments are less-structured than urban environments due to the natural randomness in the distribution of off-road features; and the considerable effort required to collect, curate and label off-road datasets makes domain adaptability a key concern. There is considerable benefits to be gained if a CNN trained with one dataset can be deployed in varied off-road environments with little modification.

In the work presented by this paper, we experimentally study the effectiveness of multi-modal fusion of RGB-LBP inputs for improved segmentation performance in challenging off-road datasets. Secondly, we demonstrate domain adaptability limitations in off-road environments by comparing the performance of networks trained on our north American off-road dataset against the Freiburg Forest dataset.

II. RELATED WORK

A. Off-Road Datasets and Challenges

Terrestrial off-road terrains can be forests with trails, or the more diverse Savannah- or prairie-like terrains. Partially-labelled real-world datasets [19] and synthetic datasets [20] have been used to demonstrate the effectiveness of transfer learning to leverage large classification datasets. But it was also shown that transfer learning is no substitute for a large dataset, and that synthetic datasets must be carefully developed to avoid negative effects in transfer learning.

Although other off-road datasets exist in the literature [21], [22], one predominant publicly available off-road dataset is the Freiburg Forest dataset [6]. However, one critical problem with this dataset is that the images are not of the same size, suggesting inconsistencies in its pre-processing. Aspect ratio is important in CNNs, particularly for smaller features such as those in the Object class. For example, images in benchmark datasets such as Cityscapes [7] are of uniform size. From Fig. 2, an important question arises: What input size to use in the CNN? In smaller networks [23], a smaller input size not only improves resolution in the encoder, it also mitigates the effects of differing aspect ratios. But deeper networks such as DeepLab [10] require larger and consistent input sizes. In addition, it is unclear why, unlike in the train set, there is no tree sample in the test set. In summary, the dataset is not reasonably uniform enough to allow for full and consistent evaluation. This is one significant motivation for creating our dataset.
We contend that our own Southern Ontario Off-Road (SOOR) dataset (Fig. 3) is more challenging due to its diversity. In addition to classes in the Freiburg Forest dataset, our SOOR dataset comprises three vegetation subdivisions: Traversable grass, Grass and Vegetation; representative of three levels of difficulty of traversal. It also includes a Water class. The SOOR dataset consists of 257 densely labelled images of size $768 \times 384$ pixels taken on three separate days in July in southern Ontario, Canada. Both datasets exhibit natural class and sample imbalance as shown in Figs. 1 and 3.

B. Multi-Modal Fusion for Improved Scene Understanding

In [6], multimodal fusion of RGB, depth and near-infrared (NIR) data was studied, and in [21] RGB and LiDAR data fusion was studied. Although fusion with 3D data can improve segmentation performance, the authors reported that the size of the fused network can limit capacity for deployment online. Also, because of the characteristic noise, stereo disparity maps required refining, and predicting depth with a separate network [24] involves an extra step. It must be noted that, unlike texture, some of these modalities, such as depth and vegetation indices, are not intrinsic properties of all the features in the off-road environment. The later the stage of fusion the more flexible the network is to changes in input modalities. Early or mid-stage fusion is achieved by adding or concatenating the weights or inputs. Late-stage fusion can also be achieved by ensembling or by mixture of experts, which is also applicable to mid-stage fusion. Practically, the most appropriate approach is ascertained by experimentation. A review of multi-modal fusion is presented in [25]. Although late-stage fusion gave the best results in [6], in our work, we use RGB-LBP channel concatenation because it only marginally increases the size of the network and makes for faster inference.

C. Texture Features

Texture features include Gabor, Haralick, wavelet, and LBP descriptors; of these, the LBP descriptor is often preferred because of its relative invariance to illumination and rotation [17]. Besides ease of implementation, another advantage of the LBP is its flexibility which has been leveraged for multi-resolution analysis [8]. The LBP patterns can be formed in the image plane, thus they fit contemporary end-to-end CNN-based navigation schemes.

D. Domain Adaptation

Here, domain adaptation refers to the flexibility to use a CNN, trained on one off-road environment, with images from another off-road environment without labelling. Domain adaptation is commonly treated as a traversability or obstacle avoidance problem [26]–[28] but, practically, in vehicle navigation, obstacle avoidance has long been implemented using 3D sensors [1]. With contemporary semantic segmentation schemes, domain adaptation is the more challenging task of finding correspondences between images at pixel level [29], [30], or as defined in [31]. Here, we experimentally demonstrate the domain adaptation challenge in two off-road environments by cross-comparing CNNs trained on the SOOR and Freiburg Forest datasets.

III. EXPERIMENTAL APPROACH

A. The multi-resolution LBP Texture Descriptor

As shown in Fig. 4, the LBP descriptor uses a binary representation of pixel neighbourhoods to define the texture of image features. For pixel $c$ with intensity $g$ in the grayscale image with a circularly symmetric number of neighbor pixels $P$ within radius $R$, the LBP is defined as [16]

$$LBP_{P,R} = \sum_{p=0}^{P-1} 2^p s(x),$$

where

$$s(x) = \begin{cases} 1 & x \geq 0 \\ 0 & x < 0, \end{cases}$$

$$x = g_p - g_c.$$  

We set $P$ as eight to correspond to the 8-bit images for uniformity during normalization. For multi-resolution analysis, multiple $R$-scale features are combined [18], but we use single $R$ descriptors per fusion. To reasonably capture local information, we limited $R$ to 3, 5 and 7 (see Fig. 4).

B. Implementation, Training and, Augmentation

Augmentation was done while training using a data generator. Positional augmentation was limited to random vertical and horizontal flips because rotations through acute angles alter labels. Vertical flips improved performance particularly for the irregular shaped water puddles. Colour augmentation was limited to brightening in the range $[0.5, \ldots, 1.5]$. Since brightening affects intensities, it was not done for the labels.

For the Freiburg Forest dataset, we used their validation set for testing. But, during training, their training set was split into the $75−25$ training-validation ratio for early stopping to limit overfitting by monitoring the validation loss profile relative to the training loss profile. And, because the dataset is small, we used three separate training-validation splits to eliminate biases in each validation set. We used the same strategy for our dataset, after keeping $10\%$ of the dataset for testing. The common learning rate policy was $1 - \frac{e}{\text{Epoch}}$ of the specified initial learning rate, where $e$ is the epoch; with a batch size of four. The learning rate was extensively tuned in each case. To evaluate the techniques on their own merits, we used the basic SGD without regularization methods such as patch extraction or cropping. Our platform was Tensorflow with an NVIDIA Quadro RTX 6000 GPU.

C. Networks

We used the FCN8 [23] network architecture as a benchmark with $448 \times 224$ pixels input size, and selected from the Cityscapes leaderboard two notable multi-scale networks, PSPNet and DeepLabv3, that use different multi-scale approaches. PSPNet uses spatial pyramid pooling while DeepLabv3 uses dilated convolutions. The implementations
IV. RESULTS

A. Results from FCN8

Results from the SOOR dataset test set are shown in Table I for two different 75-25 training-validation splits. The metric is the practical intersection over union (IOU) [32] that measures spatial matching between labels and predictions. The difficult classes are the Object, Water and Traversable grass class.

While water puddles in potholes can have similar shapes, they come in different sizes and forms. And, as shown in Fig. 3, the Water class is the most imbalanced of these, thus it is highlighted in Table I. Water can be uniquely difficult because it reflects other classes in the environment, but this reflection serves to highlight the role of texture in CNNs. As seen in the third figure in Fig. 5, in which the water puddle reflects grass, sky, etc., texture provides added discriminating information to the CNN. The texture network labels a section of the water puddle Sky according to the sky reflection on the water, but the RGB-only network struggles with this distinction. However, the multi-modal RGB-LBP network combines both information for improved segmentation.

Importantly, the significance of texture and shape to CNNs is evident in the texture-only networks in Table I. Two important cases are when \( R = 1 \) and \( R = 200 \). In the former case, the texture features are derived from pixels in the immediate neighbourhood, thus the patterns are not uniquely defined for each class as when \( R = 3, 5 \) or \( 7 \) (Fig. 4). In the latter case, the texture features are scrambled as they are derived from remote pixels. However, the overall shape of the classes as defined by their respective boundaries in the labelled images provides enough information for reasonable segmentation in the larger classes.

The interactions between traversable grass and the morphologically similar grass are seen in the wrongly segmented sections in Fig. 5. The added texture information does not significantly ameliorate the eclectic nature of the Object class. Hence the relatively poorer segmentation performance in these classes relative to sky and trail. Critically, RGB-texture fusion does not worsen overall segmentation performance.

Results from the Freiburg Forest dataset are shown in Table II. The most difficult class in this dataset is the Objects class (Fig. 1) for the same reasons as in the SOOR dataset. Merging the Tree and Vegetation classes, improves results particularly in the Object class [6]. But we considered the more challenging case by training the CNN with separate Tree and Vegetation classes. However, the Tree class is not represented in their test set, hence the overall performance is not fully evaluated. Although the improvement offered by the multi-modal RGB-LBP input over RGB-only input for the Object class is not marked, it is not insignificant.
<table>
<thead>
<tr>
<th>Modality</th>
<th>C1</th>
<th>C2</th>
<th>C3</th>
<th>C4</th>
<th>C5</th>
<th>C6</th>
<th>C7</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>LBP 8,1</td>
<td>0.818</td>
<td>0.676</td>
<td>0.674</td>
<td>0.789</td>
<td>0.386</td>
<td>0.238</td>
<td>0.156</td>
<td>0.534</td>
</tr>
<tr>
<td>LBP 8,3</td>
<td>0.854</td>
<td>0.740</td>
<td>0.740</td>
<td>0.854</td>
<td>0.493</td>
<td>0.609</td>
<td>0.155</td>
<td>0.635</td>
</tr>
<tr>
<td>LBP 8,5</td>
<td>0.866</td>
<td>0.752</td>
<td>0.747</td>
<td>0.852</td>
<td>0.487</td>
<td>0.662</td>
<td>0.711</td>
<td>0.648</td>
</tr>
<tr>
<td>LBP 8,7</td>
<td>0.870</td>
<td>0.763</td>
<td>0.721</td>
<td>0.826</td>
<td>0.475</td>
<td>0.521</td>
<td>0.198</td>
<td>0.625</td>
</tr>
<tr>
<td>LBP 8,10</td>
<td>0.870</td>
<td>0.761</td>
<td>0.694</td>
<td>0.814</td>
<td>0.425</td>
<td>0.644</td>
<td>0.191</td>
<td>0.628</td>
</tr>
<tr>
<td>LBP 8,200</td>
<td>0.797</td>
<td>0.638</td>
<td>0.604</td>
<td>0.755</td>
<td>0.308</td>
<td>0.0</td>
<td>0.016</td>
<td>0.445</td>
</tr>
<tr>
<td>RGB</td>
<td>0.908</td>
<td>0.817</td>
<td>0.806</td>
<td>0.914</td>
<td>0.609</td>
<td>0.617</td>
<td>0.350</td>
<td>0.717</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Modality</th>
<th>C1</th>
<th>C2</th>
<th>C3</th>
<th>C4</th>
<th>C5</th>
<th>C6</th>
<th>C7</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>LBP 8,1</td>
<td>0.809</td>
<td>0.647</td>
<td>0.680</td>
<td>0.814</td>
<td>0.330</td>
<td>0.016</td>
<td>0.130</td>
<td>0.489</td>
</tr>
<tr>
<td>LBP 8,3</td>
<td>0.854</td>
<td>0.739</td>
<td>0.728</td>
<td>0.835</td>
<td>0.438</td>
<td>0.543</td>
<td>0.190</td>
<td>0.618</td>
</tr>
<tr>
<td>LBP 8,5</td>
<td>0.869</td>
<td>0.753</td>
<td>0.725</td>
<td>0.835</td>
<td>0.451</td>
<td>0.653</td>
<td>0.247</td>
<td>0.648</td>
</tr>
<tr>
<td>LBP 8,7</td>
<td>0.865</td>
<td>0.758</td>
<td>0.741</td>
<td>0.849</td>
<td>0.481</td>
<td>0.600</td>
<td>0.195</td>
<td>0.641</td>
</tr>
<tr>
<td>LBP 8,10</td>
<td>0.875</td>
<td>0.758</td>
<td>0.732</td>
<td>0.851</td>
<td>0.436</td>
<td>0.595</td>
<td>0.185</td>
<td>0.633</td>
</tr>
<tr>
<td>LBP 8,200</td>
<td>0.836</td>
<td>0.647</td>
<td>0.612</td>
<td>0.827</td>
<td>0.356</td>
<td>0.093</td>
<td>0.069</td>
<td>0.492</td>
</tr>
<tr>
<td>RGB</td>
<td>0.903</td>
<td>0.806</td>
<td>0.786</td>
<td>0.904</td>
<td>0.562</td>
<td>0.609</td>
<td>0.296</td>
<td>0.695</td>
</tr>
</tbody>
</table>

1 Class: (1) Trail (2) Grass (3) Vege. (4) Sky (5) Obstacle (6) Water (7) Traversable grass.

Fig. 5. Predictions from the SOOR dataset. The discretionary basis of distinguishing Traversable grass (yellow) from Grass (green) during labelling means that the CNN often mislabels them. Also, the texture features provide added information for handling reflections on the water surface in the multi-modal network. $R$ is 5 in all cases.
TABLE II
FCN8 on Freiburg Forest dataset (IOU).

<table>
<thead>
<tr>
<th>Modality</th>
<th>C1</th>
<th>C2</th>
<th>C3</th>
<th>C4</th>
<th>C5</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>LBP₈,₃</td>
<td>0.692</td>
<td>0.766</td>
<td>0.841</td>
<td>0.879</td>
<td>0.157</td>
<td>0.556</td>
</tr>
<tr>
<td>LBP₈,₅</td>
<td>0.709</td>
<td>0.703</td>
<td>0.848</td>
<td>0.875</td>
<td>0.148</td>
<td>0.559</td>
</tr>
<tr>
<td>LBP₈,₇</td>
<td>0.720</td>
<td>0.782</td>
<td>0.840</td>
<td>0.873</td>
<td>0.166</td>
<td>0.564</td>
</tr>
<tr>
<td>LBP₈,₁₀</td>
<td>0.701</td>
<td>0.767</td>
<td>0.840</td>
<td>0.866</td>
<td>0.099</td>
<td>0.545</td>
</tr>
<tr>
<td>RGB</td>
<td>0.846</td>
<td>0.857</td>
<td>0.881</td>
<td>0.910</td>
<td>0.200</td>
<td>0.616</td>
</tr>
<tr>
<td>RGB-LBP₈,₃</td>
<td>0.849</td>
<td>0.864</td>
<td>0.868</td>
<td>0.910</td>
<td>0.196</td>
<td>0.615</td>
</tr>
<tr>
<td>RGB-LBP₈,₅</td>
<td>0.838</td>
<td>0.858</td>
<td>0.865</td>
<td>0.912</td>
<td>0.282</td>
<td>0.626</td>
</tr>
<tr>
<td>RGB-LBP₈,₇</td>
<td>0.850</td>
<td>0.862</td>
<td>0.874</td>
<td>0.907</td>
<td>0.274</td>
<td>0.623</td>
</tr>
</tbody>
</table>

1 Class: (1) Trail (2) Grass (3) Vegetation (4) Sky (5) Obstacle.

TABLE III
Multi-scale networks on SOOR dataset (IOU).

<table>
<thead>
<tr>
<th></th>
<th>FCN8</th>
<th>PSPNet</th>
<th>DeepLabv3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trail</td>
<td>0.907</td>
<td>0.907</td>
<td>0.892 0.906</td>
</tr>
<tr>
<td>Grass</td>
<td>0.814</td>
<td>0.819</td>
<td>0.797 0.814</td>
</tr>
<tr>
<td>Vegetation</td>
<td>0.775 0.794</td>
<td>0.732 0.749</td>
<td>0.799 0.804 0.914</td>
</tr>
<tr>
<td>Sky</td>
<td>0.920</td>
<td>0.910</td>
<td>0.840 0.892</td>
</tr>
<tr>
<td>Obstacle</td>
<td>0.573</td>
<td>0.601</td>
<td>0.541 0.536</td>
</tr>
<tr>
<td>Water</td>
<td>0.622</td>
<td>0.717</td>
<td>0.455 0.738</td>
</tr>
<tr>
<td>Trav. grass</td>
<td>0.336 0.330</td>
<td>0.186 0.231</td>
<td>0.307 0.318</td>
</tr>
<tr>
<td>Mean</td>
<td>0.707</td>
<td>0.726</td>
<td>0.635 0.695</td>
</tr>
</tbody>
</table>

1 Inputs: (1) RGB (2) RGB-LBP₈,₅

B. Multi-Scale Networks on SOOR Dataset

The results from the multi-scale networks, shown in Table III, follow the same trend as FCN8. Although, the FCN8 encoder used for PSPNet has limited capacity, and the deeper DeepLabv3 network improves overall performance in the Water class, the results follow the same trend for the Water class as in Table II. The LBP texture features do not interfere with the separate multi-scale techniques used in each network.

C. Off-Road Domain Adaptability

Domain adaptation in structured environments, with focus on the domain shift between synthetic and real images, is well-studied using adversarial learning techniques [33]–[36]. The domain shift between real images of different cities was studied with only one source dataset in [37], [38]. We contend that the domain shift in off-road environments are comparatively larger because of the natural random distribution and variation of environmental features. Also, because the textural patterns of off-road classes are relatively consistent, textural features are relevant [12], [13]. We explore off-road domain adaptation using the Freiburg Forest and SOOR datasets as source and target domains separately, and highlight the effect of texture.

For uniformity, we merged some common classes in each dataset according to Table IV. For the SOOR dataset the Grass and Traversable grass classes were merged; and for the Freiburg Forest dataset, the Tree and Vegetation classes were merged. Images with water were not used in the test. Sample images from each dataset are shown in Fig. 6. In our tests we trained four DeepLabv3 networks on each dataset with RGB-only and RGB-LBP input modalities, respectively.

For the networks trained and evaluated on the same dataset, merging similar classes made each class more distinct and reduced the role of texture as seen in the comparative segmentation performance of both input modalities in Table V. A different fusion strategy [6], [25] or choice of LBP resolution can give different results.

However, the results are different for networks evaluated on a different dataset. As shown in Fig. 6, the Trail, Vegetation and Obstacle classes are the most different across both environments. While networks trained on the higher resolution SOOR dataset gave better results, overall, the best results are for the Sky class, which is the most consistent amongst the classes across both environments. The texture-fused networks reflect this consistency, while emphasizing the innate differences in the Grass class, which is less uniform in the SOOR dataset (Fig. 6).
TABLE V  
MULTI-DOMAIN TESTING RESULTS (IOU).

<table>
<thead>
<tr>
<th>Input</th>
<th>C1</th>
<th>C2</th>
<th>C3</th>
<th>C4</th>
<th>C5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Freiburg Forest on Freiburg Forest</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>In1</td>
<td>0.855</td>
<td>0.862</td>
<td>0.889</td>
<td>0.911</td>
<td>0.328</td>
</tr>
<tr>
<td>In2</td>
<td>0.826</td>
<td>0.842</td>
<td>0.880</td>
<td>0.907</td>
<td>0.325</td>
</tr>
<tr>
<td>Freiburg Forest on SOOR</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>In1</td>
<td>0.007</td>
<td>0.220</td>
<td>0.100</td>
<td>0.289</td>
<td>0.001</td>
</tr>
<tr>
<td>In2</td>
<td>0.074</td>
<td>0.048</td>
<td>0.121</td>
<td>0.460</td>
<td>0.000</td>
</tr>
<tr>
<td>SOOR on SOOR</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>In1</td>
<td>0.890</td>
<td>0.813</td>
<td>0.795</td>
<td>0.920</td>
<td>0.566</td>
</tr>
<tr>
<td>In2</td>
<td>0.893</td>
<td>0.810</td>
<td>0.800</td>
<td>0.906</td>
<td>0.572</td>
</tr>
</tbody>
</table>

V. CONCLUSION

We have shown that CNNs are sensitive to LBP texture features at different resolutions, and that fusing RGB and LBP texture features can improve segmentation performance compared to RGB-only segmentation in imbalanced off-road datasets. Like any hyper-parameter, texture parameters, and the fusion strategy, must be tuned by extensive experimentation. Texture features play a role particularly in difficult off-road datasets that have some non-unique classes. Also, we have experimentally demonstrated the domain adaptation problem in off-road environments and shown the importance of texture features across off-road environments.

In order to reduce the tuning effort, we plan to formulate a method for determining the best LBP parameters based on the image parameters. In addition, we hope to experiment with late fusion strategies that make the effect of individual input modalities more pronounced. Finally, we plan to implement a domain adaptation network on both datasets.

REFERENCES


