7 Supplement

7.1 Limitations

We term our method hybrid since it optimizes two different training objectives: i) dense discrimination between inliers and negatives, and ii) high likelihood of inliers and low likelihood of negative data. It may seem that our method can generate samples due to likelihood evaluation being a standard feature of generative models (except GANs). However, our formulation is not suitable for sample generation due to dealing with unnormalized distributions. This would require MCMC sampling which can not be performed at large resolutions, at least not with known techniques. Even if sample generation was feasible, the resulting approach would likely be too slow for real-time inference as shown by other image resynthesis approaches [36,4].

7.2 Impact of Known Inliers on Anomaly Detection

Many real-world deployments of autonomous systems work in environment with limited variety (e.g. warehouses or industrial plants). Such environments usually have perfectly aligned training and test distributions. Still, anomalous objects can occur. We show that anomaly detection performance in such cases is significantly more easier. Table 7 shows such setup. We used two DLv3+ segmentation networks. The first one is trained on Cityscapes train while the second one is trained on Cityscapes train and val splits. Since the Fishyscapes Static dataset is created by pasting negative objects atop Cityscapes val images, we can measure anomaly detection performance when the inlier instances from test set are known. We see that the average precision of anomaly detection is drastically improved from 60% to 89%. This indicates that the DenseHybrid anomaly detector is feasible for scenarios with limited scene variety.

Table 7. Impact of known inlier instances on anomaly detection performance. Results indicate that DenseHybrid is feasible for scenarios with limited scene variety

Training splits	FS Static		Closed world
	AP	FPR_{95}	Cityscapes val mIoU
train	60.0±2.0	4.9±0.6	81.0
train+val	88.5±0.8	1.1±0.1	89.9

7.3 More results

Table 8 shows the dense open-set recognition performance with generative and discriminative anomaly detectors on the StreetHazards dataset. The proposed hybrid anomaly score based on the ratio of these two distributions outperforms each of the two components. Note that generative and discriminative outputs are trained using the same mixed-content images.

Table 8. Performance evaluation on StreetHazards. DenseHybrid anomaly score based on the ratio of generative and discriminative distributions outperforms each of the two components

Method	Anomaly detection			Closed world			Open world		
	AP	FPR_{95}	AUC	$\overline{\text{IoU}}$-t5	$\overline{\text{IoU}}$-t6	$\overline{\text{IoU}}$	o-$\overline{\text{IoU}}$-t5	o-$\overline{\text{IoU}}$-t6	o-$\overline{\text{IoU}}$
Generative	30.0	13.3	95.5	65.6	61.6	63.0	45.6	45.2	45.5
Discriminative	23.3	20.5	93.1	65.6	61.6	63.0	36.9	35.2	36.4
DenseHybrid (ours)	**30.2**	**13.0**	**95.6**	65.6	61.6	63.0	**46.1**	**45.3**	**45.8**

7.4 Implementation Details

We create mixed content training samples by pasting negative data atop the inlier crops. The inlier crops are obtained by jittering images in range [0.5, 2], applying random horizontal flip, and taking random square crop of size 768. We use ADE20k instances as negative content. We resize the negative image to 768 pixels, take a random jittered crop of size 384. Then, we paste two instances from the negative crop at random position atop the inlier crop. For LDN-121, we use batch size 16. We use Adam optimizer with the initial learning rate 10^{-5}. The learning rate is decayed through a cosine schedule down to 10^{-7}. We set the loss modulation hyper-parameter β to 0.03. For DLv3+ we use batch size 8 and inlier crops of 512. We use Adam optimizer with learning rate 10^{-6} and do not decay it. We set the loss modulation hyper-parameter β to 0.01.

In the case of Fishyscapes, we fine-tune DLv3+ with a WRN38 backbone pre-trained by NVIDIA [55]. However, due to hardware limitations we could not train it from scratch to achieve the desired robustness required for SMIYC. Hence, we opted for LDN-121 as an efficient alternative which can be trained on a single GPU. Using the bigger DLv3+ would additionally improve results on SMIYC.

7.5 Visualizations

Figure 7 visualizes anomaly detection performance of DenseHybrid on SMIYC-AnomalyTrack and SMIYC-ObstacleTrack. Anomalies detected with DenseHybrid are highlighted above the input image. The corresponding ground-truth has anomalies designated in orange. Inlier pixels are designated in white, while the ignore pixels are designated in black.

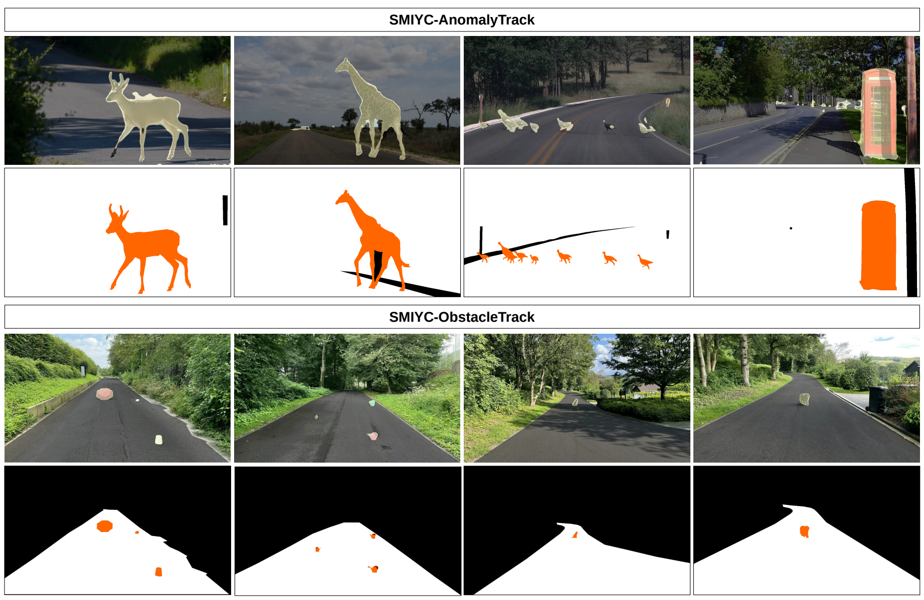

Fig. 7. Anomaly detection performance of DenseHybrid on SMIYC validation subsets