Wormhole Learning

Alessandro Zanardi[1], Julian Zilly[1], Andreas Aumiller[1], Andrea Censi[1], Emilio Frazzoli[1]
[1]**Institute for Dynamic Systems and Control, ETH Zurich**

1 Introduction

Typically, **to enlarge** the operating **domain of an object detector**, more labeled training data is required. We introduce a novel method called Wormhole Learning, which allows to extend the operating domain without additional labeled data, but **only with temporary access to an auxiliary sensor** with certain invariance properties. We showcase the instantiation of this principle with a regular visible-light **RGB camera as the main sensor**, and an **infrared sensor as the temporary auxiliary sensor**.

2 Method overview

Wormhole Learning can be generalized into three steps. We first start with a RGB detector **pre-trained on daytime data only**; given a stream of paired sensor data, we then train the infrared detector based on the RGB-inferred labels. In a second step, we **exploit the inherent invariance** of the infrared sensor to scaling of ambient illumination and are thus able to **infer labels at night**. In the last step, the sensors switch roles and we perform **transfer learning back** to the RGB domain. The re-trained RGB detector now has **enlarged its operating domain by partly inheriting the auxiliary sensor's invariance** to illumination; in particular, the RGB detector is able to perform much better at difficult lightning situation as such at night.

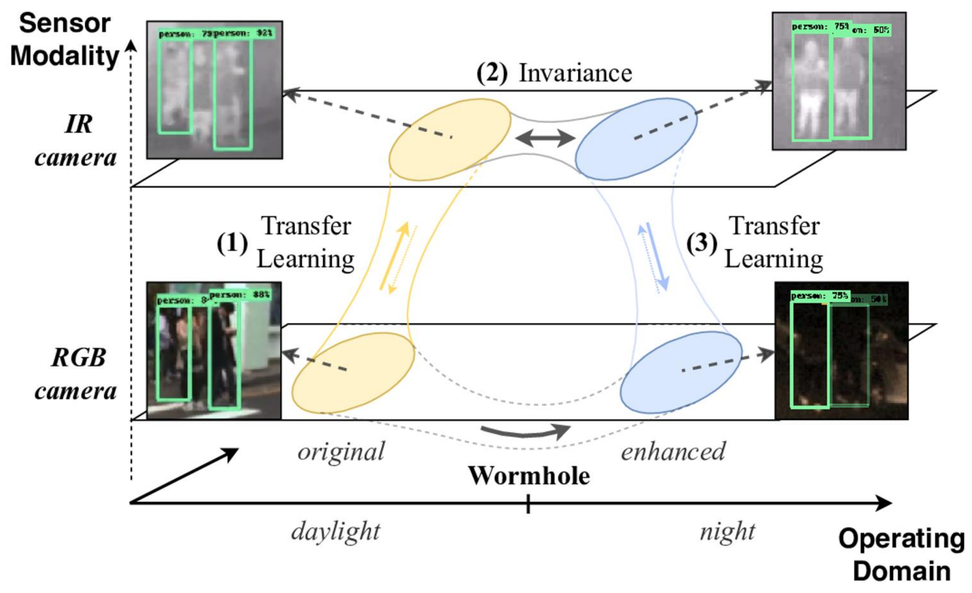

Fig. 1. A wormhole can be created in the operating domain of a sensor leveraging the inherent invariance of another auxiliary sensor—in this case invariance to illumination of an IR camera. Starting with an object detector trained only with daytime data we enlarge its capabilities to also include night time thanks to the temporary addition of an IR camera. The night-day wormhole is created from three steps:
1) Transfer learning from RGB camera to IR at daytime.
2) Exploiting the invariance of IR camera with respect to time of the day.
3) Transfer learning back from IR to RGB at night.

3 Technical Insights

We define the **wormhole gain** as the difference in cross-entropy for the detector *before* ($q_{\theta^{D}_{RGB}}$) and *after* ($q_{\theta^{D+N}_{RGB}}$) the wormhole learning with respect to an unobservable underlying distribution: the ground-truth relative to the scene ($p_{\bar{Y}}$)

$$\mathrm{WG}^{D+N}_{RGB\rightarrow IR} = \mathbb{E}_{\mathcal{D}^{D+N}}\left[H\left(p_{\bar{Y}}, q_{\theta^{D}_{RGB}}(\mathrm{Y}|\mathrm{Z}_{RGB})\right) - H\left(p_{\bar{Y}}, q_{\theta^{D+N}_{RGB}}(\mathrm{Y}|\mathrm{Z}_{RGB})\right)\right],$$

where as Y denotes the inferred label by an object detector parameterized as θ and Z depicts the representation of a scene as captured in one sensor's modality.
Moreover, defining a Jaccard **similarity index** for two sensors as

$$J_{RGB,IR} = \frac{I(Z_{RGB}; Z_{IR})}{H(Z_{RGB}, Z_{IR})}, \qquad 0 \leq J_{RGB,IR} \leq 1$$

we show that wormhole learning to be successful requires the sensors to be **neither «*too orthogonal*»** ($J_{RGB,IR} = 0$) **nor «*too similar*»** ($J_{RGB,IR} = 1$). The interested reader is kindly referred to [2] for the details.

4 Results and discussion

Fig. 2. After wormhole learning (*right*) we learned to recognize cars only from the headlights.

We empirically validate the concept of wormhole learning in an experiment on the KAIST multi-spectral dataset [1]. A synchronous stream of RGB and IR images is provided along a split between daytime (D) and night (N) data. A pre-trained Faster-RCNN network [3] using the NASnet architecture [4] is employed in order to generate ground-truth data for the initial training set at daytime. Furthermore, for each **domain transfer** step we started from the same checkpoint of a single shot detector architecture pre-trained on the COCO [5] dataset. We included 6 object classes in our detection task, namely c*ar, person, bus, truck, motorcycle* and *bicycle.*

Table 1
Detection performance in mAP@0.5IoU

Testset environment	RGB (D)	IR	RGB (D+N)	Relative Gain
day	43.7	22.4	41.6	**-5.0%**
night	13.4	31.2	20.3	**+51.2%**

We observe in Table 1, that the IR detector trained using labels inferred from the original RGB detector performs worse at daytime, but excels at night due to its approximate invariance to ambient lighting. The 4th column shows the effect of **wormhole learning** as the re-trained RGB detector has gained a huge relative performance boost at night, while operating slightly worse at day.

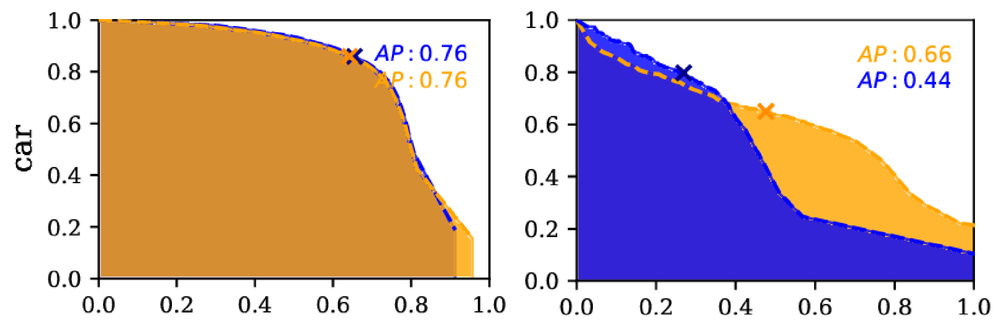

Fig. 3. Recall-precision curves for category *car*. In **blue** the detector before WHL, in **orange** the detector after WHL. We can appreciate a significant improvement in performance at **night (right panel)**, while maintaining comparable performance during **day (left panel)**. In particular, one could speculate that we increased performance at high recall because we learned a whole new set of representations for the object category (cf. Fig. 3).

5 Conclusion

We introduced **Wormhole Learning** as a novel way of leveraging the mutual information between a main and an auxiliary sensor to enlarge the operating domain of the former via semi-supervised learning. Furthermore, this provides a simple way of generating "unlimited" labeled data at no cost. Crucially, we showed that **invariance to undesired changes in data** of the auxiliary sensor **can be exploited to improve learning outcomes** for the first sensor.

6 References

1. Y. Choi, N. Kim, S. Hwang, K. Park, J. S. Yoon, K. An, and I. S. Kweon, "KAIST Multi-Spectral Day/Night Data Set for Autonomous and Assisted Driving," IEEE Transactions on Intelligent Transportation Systems, 2018.
2. A. Zanardi, J. Zilly, A. Aumiller, A. Censi, E. Frazzoli, "Wormhole learning," IEEE International Conference on Robotics and Automation (ICRA), 2019.
3. Jonathan Huang, Vivek Rathod, Chen Sun, Menglong Zhu, Anoop Korattikara, Alireza Fathi, Ian Fischer, Zbigniew Wojna, Yang Song, Sergio Guadarrama, Kevin Murphy, "Speed/accuracy trade-offs for modern convolutional object detectors, " IEEE Conference on Computer Vision and Pattern Recognition, 2017
4. B. Zoph, V. Vasudevan, J. Shlens, and Q. V. Le, "Learning transferable architectures for scalable image recognition," arXiv preprint arXiv:1707.07012, vol. 2, no. 6, 2017.
5. T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick, "Microsoft coco: Common objects in context," in European conference on computer vision. Springer, 2014, pp. 740–755.