MLSL: Multi-Level Self-Supervised Learning for Domain Adaptation with Spatially Independent and Semantically Consistent Labeling

Javed Iqbal and Mohsen Ali
Information Technology University, Pakistan.
(javed.iqbal, mohsen.ali)@itu.edu.pk

In this document, we provide additional materials to supplement our main submission. We show additional qualitative results on Cityscapes validation set especially for small and less-occurring objects, e.g., rider, bicycle, bus etc. Fig. 1 and Fig. 3 illustrates such cases for *GTA-V to Cityscapes* and *SYNTHIA to Cityscapes* respectively. Similarly, in Fig. 2 and 4, we provide a qualitative comparison with [1] and [2], and also provide error maps for each segmentation output. To highlight the performance gap with fully-supervised (Oracle) baseline models, we trained ResNet-38 in a fully-supervised manner on Cityscapes training set having 2-images per mini-batch. Table. 1 and Table. 2 provides a detailed performance gap comparison of oracle and adaptation method, for proposed and existing state-of-the-art methods.

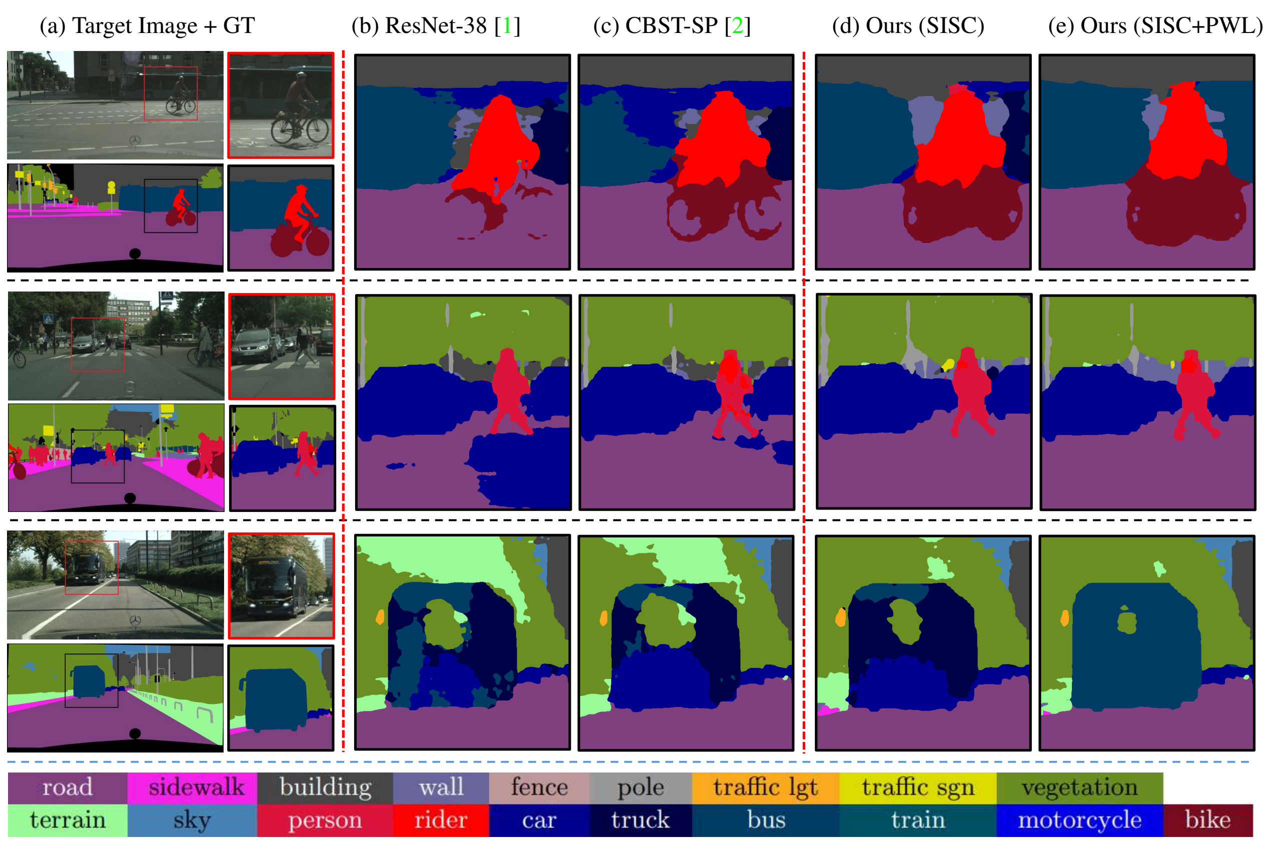

Figure 1. Qualitative results for GTA-V → Cityscapes adaptation. Column (a) (top) shows target image and a selected sub-image along with corresponding segmentation ground truth labels of sub-image and whole image (bottom) respectively. Column (b), (c), (d) and (e) shows the segmentation results for selected sub-image produced by different approaches.

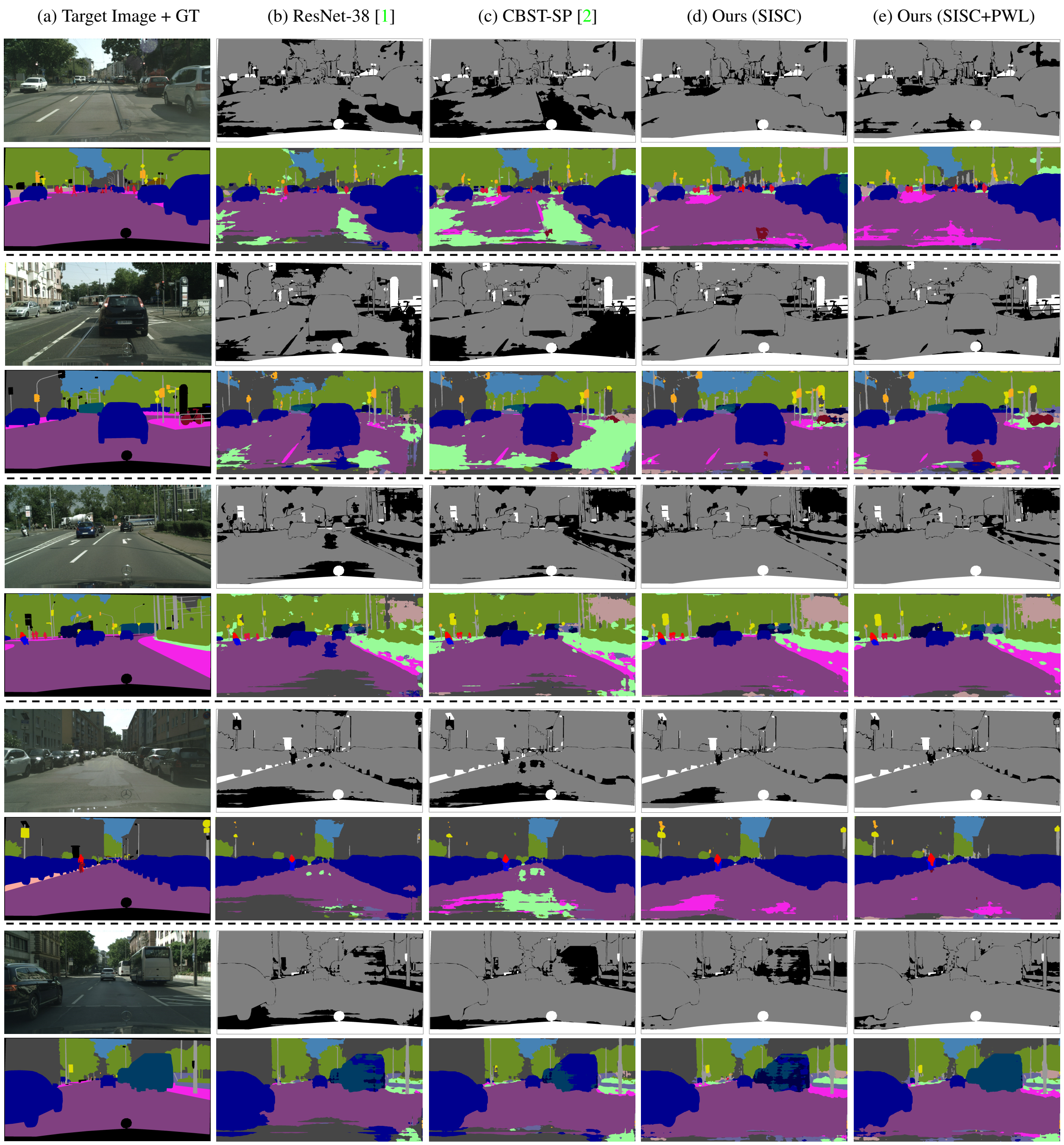

Figure 2. Qualitative results for GTA-V → Cityscapes adaptation. Column (a) shows target image along with corresponding segmentation ground truth labels. Column (b), (c), (d) and (e) shows the segmentation results (bottom) and corresponding error maps produced by different approaches (top). In each error map, the black region shows the segmentation error, white region shows the don't care region and they grey region as the correctly segmented region respectively.

Table 1. Performance (mIoU) gap between the fully-supervised (Oracle) models and adapted models from GTA-V to Cityscapes. Despite high Oracle limits, the proposed UDA methods have lower performance gap compared to other state-of-the-art methods.

GTA $\rightarrow$ Cityscapes			
Methods	Oracle	UDA Algo.	mIoU gap
FCN in the wild [3]	64.6	27.1	-37.5
Curriculam DA [4]	60.3	28.9	-31.4
AdaptSetNet [5]	65.1	42.4	-22.7
Saleh et al [6]	65.1	42.5	-22.6
MinEnt [7]	65.1	42.3	-22.8
CLAN [8]	65.1	43.2	-21.9
All Structure [9]	65.1	45.4	-19.7
CBST-SP [2]	67.6	46.2	-21.4
Ours (SISC)	67.6	48.7	-18.9
Ours (SISC+PWL)	67.6	**49.0**	**-18.6**

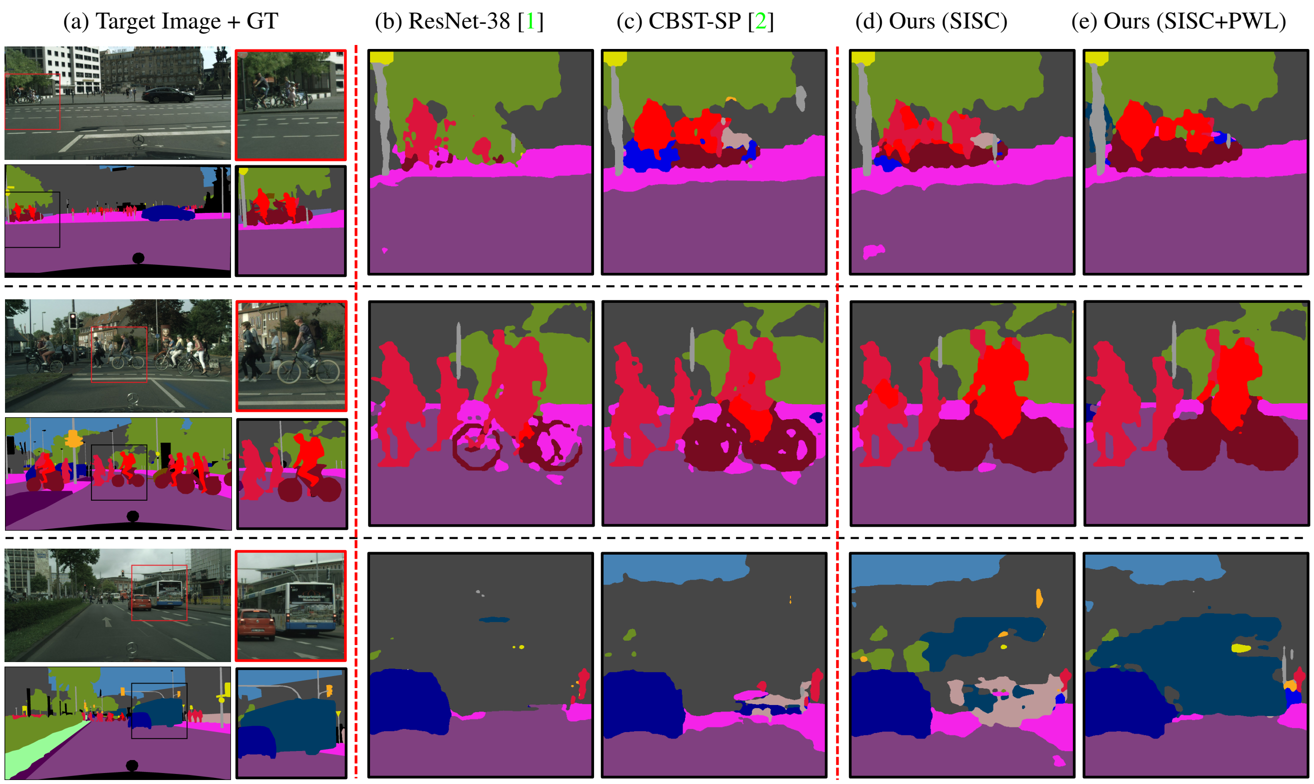

Figure 3. Qualitative results for SYNTHIA $\rightarrow$ Cityscapes adaptation. Column (a) (top) shows target image and a selected sub-image along with corresponding segmentation ground truth labels of sub-image and whole image (bottom) respectively. Column (b), (c), (d) and (e) shows the segmentation results for selected sub-image produced by different approaches.

Table 2. Performance (mIoU* (13 common classes)) gap between the fully-supervised (Oracle) models and adapted models from SYNTHIA to Cityscapes.

SYNTHIA $\rightarrow$ Cityscapes			
Methods	Oracle	UDA Algo.	mIoU* gap
FCN in the wild [3]	73.8	22.9	-50.9
Curriculam DA [4]	69.6	34.8	-34.8
AdaptSetNet [5]	71.7	46.7	-25.0
MinEnt [7]	71.7	44.2	-27.5
CLAN [8]	71.7	47.8	-23.9
All Structure [9]	71.7	48.7	-23.0
CBST [2]	73.7	48.4	-25.3
Ours (SISC)	73.7	50.8	-22.9
Ours (SISC+PWL)	73.7	**51.0**	**-22.7**

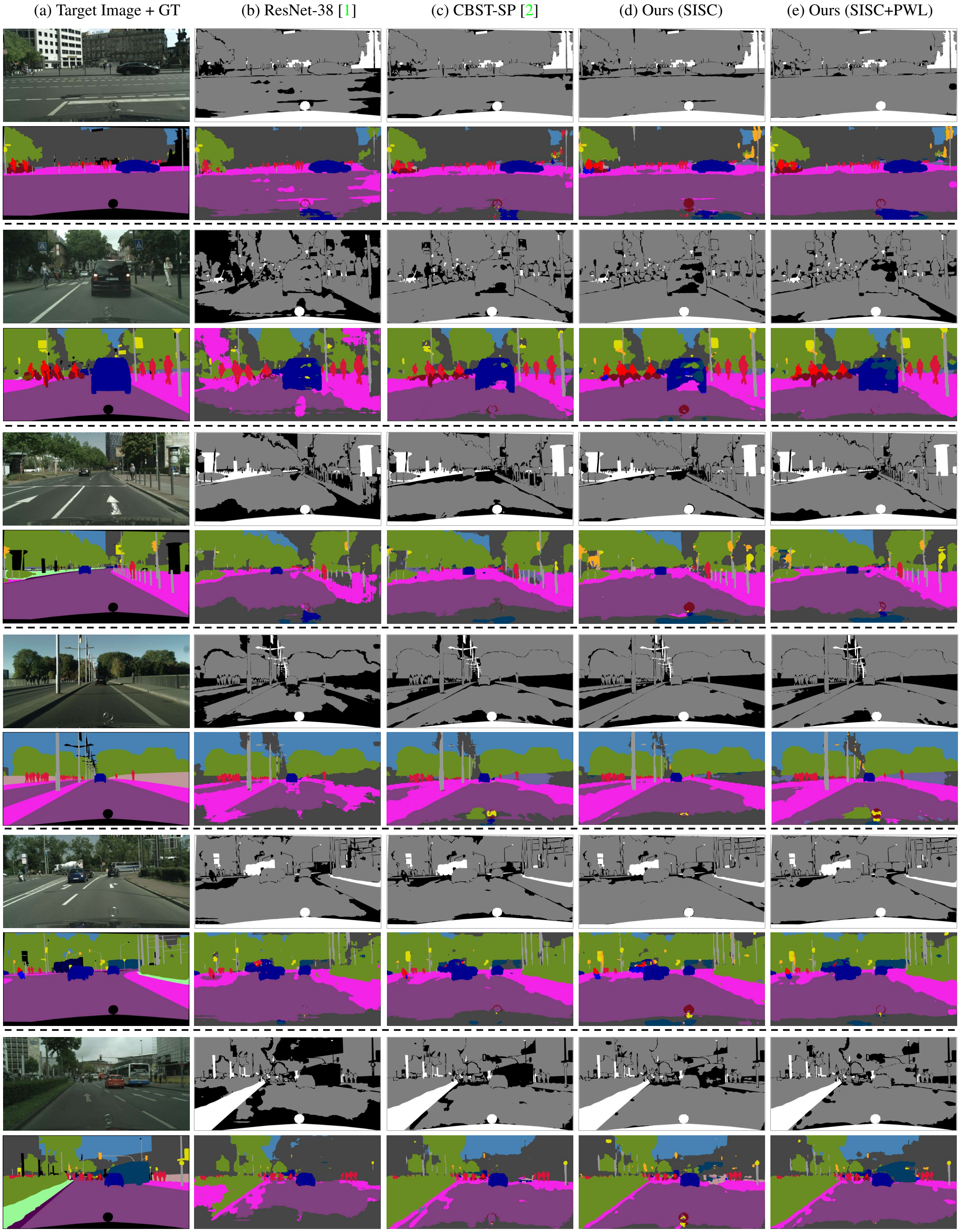

Figure 4. Qualitative results for SYNTHIA → Cityscapes adaptation. Column (a) shows target image along with corresponding segmentation ground truth labels. Column (b), (c), (d) and (e) shows the segmentation results (bottom) and corresponding error maps produced by different approaches (top). In each error map, the black region shows the segmentation error, white region shows the don't care region and they grey region as the correctly segmented region respectively.

References

[1] Zifeng Wu, Chunhua Shen, and Anton Van Den Hengel. Wider or deeper: Revisiting the resnet model for visual recognition. *Pattern Recognition*, 90:119–133, 2019. 1, 2, 3, 4

[2] Yang Zou, Zhiding Yu, BVK Vijaya Kumar, and Jinsong Wang. Unsupervised domain adaptation for semantic segmentation via class-balanced self-training. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 289–305, 2018. 1, 2, 3, 4

[3] Judy Hoffman, Dequan Wang, Fisher Yu, and Trevor Darrell. Fcns in the wild: Pixel-level adversarial and constraint-based adaptation. *arXiv preprint arXiv:1612.02649*, 2016. 3

[4] Yang Zhang, Philip David, and Boqing Gong. Curriculum domain adaptation for semantic segmentation of urban scenes. In *The IEEE International Conference on Computer Vision (ICCV)*, Oct 2017. 3

[5] Yi-Hsuan Tsai, Wei-Chih Hung, Samuel Schulter, Kihyuk Sohn, Ming-Hsuan Yang, and Manmohan Chandraker. Learning to adapt structured output space for semantic segmentation. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2018. 3

[6] Fatemeh Sadat Saleh, Mohammad Sadegh Aliakbarian, Mathieu Salzmann, Lars Petersson, and Jose M Alvarez. Effective use of synthetic data for urban scene semantic segmentation. In *European Conference on Computer Vision*, pages 86–103. Springer, 2018. 3

[7] Tuan-Hung Vu, Himalaya Jain, Maxime Bucher, Matthieu Cord, and Patrick Pérez. Advent: Adversarial entropy minimization for domain adaptation in semantic segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2517–2526, 2019. 3

[8] Yawei Luo, Liang Zheng, Tao Guan, Junqing Yu, and Yi Yang. Taking a closer look at domain shift: Category-level adversaries for semantics consistent domain adaptation. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2019. 3

[9] Wei-Lun Chang, Hui-Po Wang, Wen-Hsiao Peng, and Wei-Chen Chiu. All about structure: Adapting structural information across domains for boosting semantic segmentation. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2019. 3