Guided Curriculum Model Adaptation and Uncertainty-Aware Evaluation for Semantic Nighttime Image Segmentation – Supplementary Material –

Christos Sakaridis[1], Dengxin Dai[1], and Luc Van Gool[1,2]

[1]ETH Zürich, [2]KU Leuven

A. Proof of Theorem 1

Proof. For brevity in the proof, we drop the class superscript (c) which is used in the statement of the theorem.

Firstly, we draw an association between pixel sets related to the standard $\text{IoU} = \text{UIoU}(1/C)$ and their counterparts for UIoU defined in equations (7)–(11) of the main paper. In particular, the following holds true:

$$\begin{aligned} &|\text{TP}(1/C)| + |\text{FN}(1/C)| \\ &= |\text{TP}(\theta)| + |\text{FN}(\theta)| + |\text{TI}(\theta)| + |\text{FI}(\theta)|, \, \forall \theta \in [1/C, \, 1]. \end{aligned} \tag{13}$$

The first assumption of Th. 1 implies that $\text{FI}(\theta_1) = \emptyset$, because $\forall \theta < \theta_2$ (including θ_1) there exists no false invalid pixel for the examined class. Thus, applying (13) for $\theta = \theta_1$ leads to

$$|\text{TP}(1/C)| = |\text{TP}(\theta_1)| + |\text{TI}(\theta_1)| + |\text{FN}(\theta_1)| - |\text{FN}(1/C)|. \tag{14}$$

Secondly, we plug the proposition of the first assumption of the theorem into the proposition of the second assumption to obtain

$$(\text{FN}(1/C) \cup \text{FP}(1/C)) \setminus (\text{FN}(\theta_1) \cup \text{FP}(\theta_1)) \neq \emptyset. \tag{15}$$

We further elaborate on (15) by observing that $\text{FN}(1/C) \cap \text{FP}(1/C) = \emptyset$, $\text{FN}(\theta_1) \subseteq \text{FN}(1/C)$ and $\text{FP}(\theta_1) \subseteq \text{FP}(1/C)$ to arrive at

$$(|\text{FN}(1/C)| - |\text{FN}(\theta_1)|) + (|\text{FP}(1/C)| - |\text{FP}(\theta_1)|) > 0. \tag{16}$$

Both terms on the left-hand side of (16) are nonnegative based on our previous observations, while at the same time (16) implies that at least one of the two is strictly positive. To complete the proof, we distinguish between the two corresponding cases.

In the first case, the first term in (16) is strictly positive, so (14) implies

$$|\text{TP}(1/C)| < |\text{TP}(\theta_1)| + |\text{TI}(\theta_1)|. \tag{17}$$

We establish the inequality we are after by writing

$$\begin{aligned} &\text{IoU} = \\ &= \frac{|\text{TP}(1/C)|}{|\text{TP}(1/C)| + |\text{FN}(1/C)| + |\text{FP}(1/C)|} \\ &= \frac{|\text{TP}(1/C)|}{|\text{TP}(\theta_1)| + |\text{FN}(\theta_1)| + |\text{TI}(\theta_1)| + |\text{FI}(\theta_1)| + |\text{FP}(1/C)|} \\ &\leq \frac{|\text{TP}(1/C)|}{|\text{TP}(\theta_1)| + |\text{TI}(\theta_1)| + |\text{FP}(\theta_1)| + |\text{FN}(\theta_1)| + |\text{FI}(\theta_1)|} \\ &< \frac{|\text{TP}(\theta_1)| + |\text{TI}(\theta_1)|}{|\text{TP}(\theta_1)| + |\text{TI}(\theta_1)| + |\text{FP}(\theta_1)| + |\text{FN}(\theta_1)| + |\text{FI}(\theta_1)|} \\ &= \text{UIoU}(\theta_1), \end{aligned} \tag{18}$$

where we have used the definition of IoU in the second line, (13) in the third line, $\text{FP}(\theta_1) \subseteq \text{FP}(1/C)$ in the fourth line, (17) in the fifth line, and the definition of UIoU that has been introduced in equation (12) of the main paper in the last line.

In the second case, the second term in (16) is strictly positive, which implies that

$$|\text{FP}(1/C)| > |\text{FP}(\theta_1)|. \tag{19}$$

Besides, applying the nonnegativity of the first term in (16) to (14) leads to

$$|\text{TP}(1/C)| \leq |\text{TP}(\theta_1)| + |\text{TI}(\theta_1)|. \tag{20}$$

Similarly to the first case, we establish the inequality we are

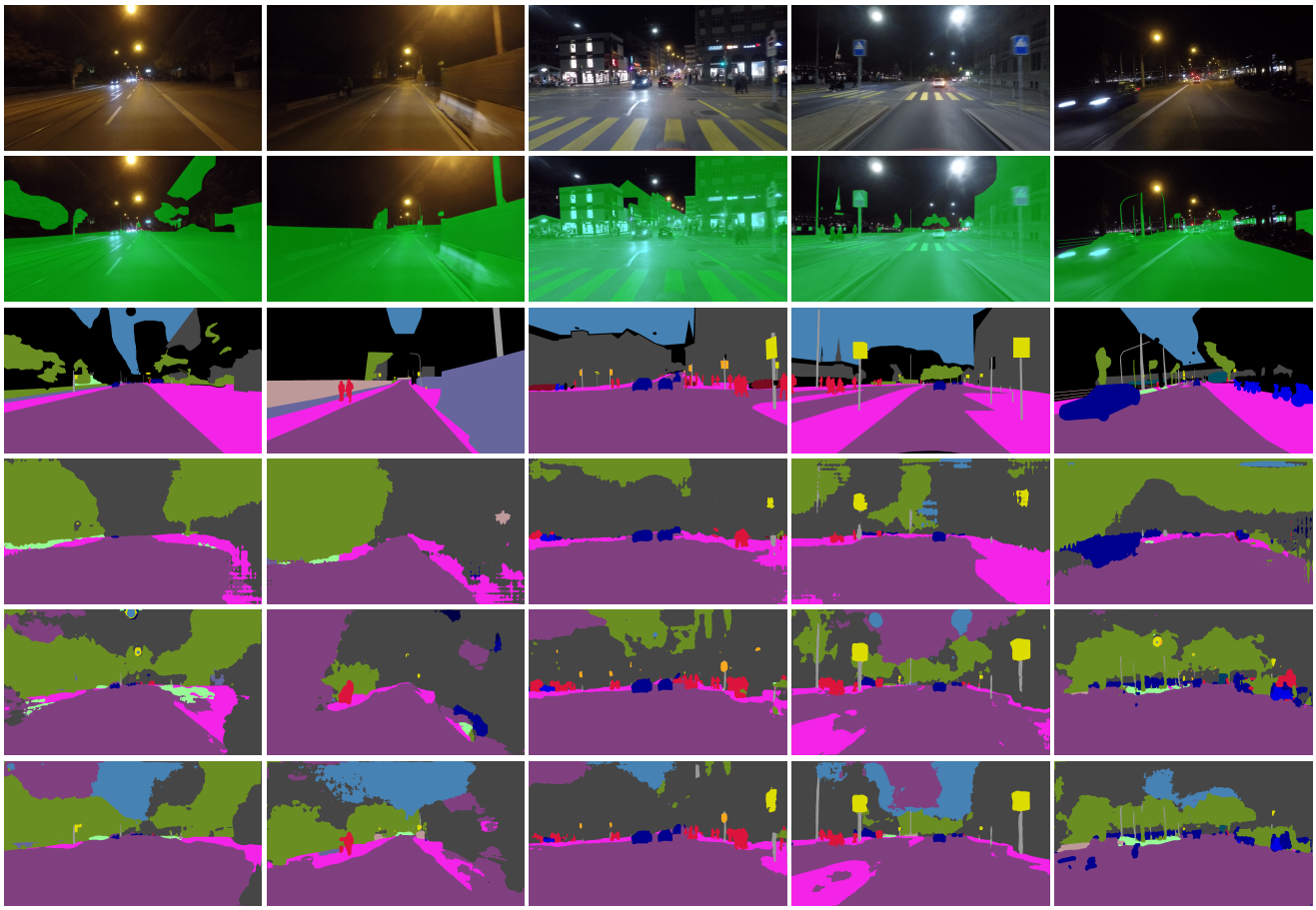

Figure 5. Examples of our annotations and qualitative semantic segmentation results on *Dark Zurich-test*. From top to bottom row: nighttime image, invalid mask annotation overlaid on the image (valid pixels are colored green), semantic annotation, AdaptSegNet [31], DMAda [8], and GCMA (ours).

after by writing

$$
\begin{aligned}
&\text{IoU} = \\
&= \frac{|\text{TP}(1/C)|}{|\text{TP}(\theta_1)| + |\text{TI}(\theta_1)| + |\text{FP}(1/C)| + |\text{FN}(\theta_1)| + |\text{FI}(\theta_1)|} \\
&< \frac{|\text{TP}(1/C)|}{|\text{TP}(\theta_1)| + |\text{TI}(\theta_1)| + |\text{FP}(\theta_1)| + |\text{FN}(\theta_1)| + |\text{FI}(\theta_1)|} \\
&\leq \frac{|\text{TP}(\theta_1)| + |\text{TI}(\theta_1)|}{|\text{TP}(\theta_1)| + |\text{TI}(\theta_1)| + |\text{FP}(\theta_1)| + |\text{FN}(\theta_1)| + |\text{FI}(\theta_1)|} \\
&= \text{UIoU}(\theta_1), \qquad (21)
\end{aligned}
$$

where we have used the definition of IoU as well as (13) in the second line, (19) in the third line, (20) in the fourth line, and the definition of UIoU in the last line. □

B. Additional Qualitative Results

In Fig. 5, we compare our GCMA approach against AdaptSegNet [31] and DMAda [8] on additional images from *Dark Zurich-test*, further demonstrating the superiority of GCMA. For these images, we also present our annotations for invalid masks and semantic labels, which show that a significant portion of ground-truth invalid regions is indeed assigned a reliable semantic label through our annotation protocol and can thus be included in the evaluation.

C. Configuration of Training Sets for GCMA

In Fig. 6, we show examples from the six training sets we introduced in Sec. 3.1, which are used for implementing GCMA. Cityscapes is used to instantiate the labeled sets, while *Dark Zurich* is used for the unlabeled sets.

More examples of Cityscapes images stylized to nighttime using a CycleGAN model [45] that is trained to translate Cityscapes to *Dark Zurich-night* are presented in Fig. 7.

D. Parameter Selection for Prediction Fusion

For our confidence-adaptive prediction fusion, we demonstrate the benefit of selecting $\alpha_l < \alpha_h < 1$—the ra-

(a) $\mathcal{D}_{lr}^{1}$: Cityscapes

(b) $\mathcal{D}_{ur}^{1}$: *Dark Zurich-day*

(c) $\mathcal{D}_{ls}^{2}$: Cityscapes-twilight style

(d) $\mathcal{D}_{ur}^{2}$: *Dark Zurich-twilight*

(e) $\mathcal{D}_{ls}^{3}$: Cityscapes-nighttime style

(f) $\mathcal{D}_{ur}^{3}$: *Dark Zurich-night*

Figure 6. Sample images from the training sets used in GCMA.

tionale of which is exposed in Sec. 3.2.2—through a visual example in Fig. 8.

Figure 7. Top row: Examples of images from Cityscapes ($\mathcal{D}_{lr}^{1}$ in GCMA), bottom row: corresponding images from Cityscapes-nighttime style ($\mathcal{D}_{ls}^{3}$ in GCMA).

(a) Dark image I^z (b) $\alpha_l = \alpha_h = 1$ (c) $\alpha_l = \alpha_h = 0.6$ (d) $\alpha_l = 0.3, \alpha_h = 0.6, \eta = 0.2$

Figure 8. Dark image I^z from *Dark Zurich* and our refined predictions $\hat{\mathbf{S}}^z$ for the region indicated by the red box for different values of the parameters involved in the proposed confidence-adaptive prediction fusion. When $\alpha_l = \alpha_h$, reducing α_h to a value lower than 1, e.g. (b)→(c), reduces false positives and/or false negatives both for static and dynamic classes, e.g. *pole*, *sidewalk*, *road* and *car*. When $\alpha_h < 1$, reducing α_l to a value lower than α_h, e.g. (c)→(d), improves accuracy on pixels that are assigned to a dynamic class in either prediction, e.g. *car*, because of the formulation of equation (6) of the main paper. Best viewed with zoom.