Supplementary Material: Increasing the Robustness of Semantic Segmentation Models with Painting-by-Numbers

Christoph Kamann[1,2] and Carsten Rother[2]

[1] Corporate Research, Robert Bosch GmbH, Renningen, Germany
`christoph.kamann@bosch.com`
[2] Visual Learning Lab, Heidelberg University (HCI/IWR), Heidelberg, Germany
`carsten.rother@iwr.uni-heidelberg.de`
http://vislearn.de

We provide further results and illustrations as mentioned in the main paper. In more detail, this supplementary material is structured as follows. We first discuss the computational efficiency of Painting-by-Numbers in section 1 and illustrate the utilized ImageNet-C corruptions, to generate Cityscapes-C, in section 2. We then discuss in section 3 in more detail the effect of several image corruptions of category *digital* and *weather*. In section 4, we list the individual mIoU scores of Table 1 of the main paper, i.e., the mIoU for each severity level per image corruption. In section 5, we extend an experiment from the main paper for understanding Painting-by-Numbers.

1 Computational Efficiency

Painting-by-Numbers is indeed efficient. We implemented Painting-by-Numbers on CPU using TensorFlow and augmented half of the images of each mini-batch on-the-fly. For our setup (a machine with 4x 1080 Ti (11 GB), and Intel Xeon CPU E5-2699 with 44 CPU cores with 2.2 GHz), the training time with Painting-by-Numbers increases by approx. 2.5 %. Therefore, it was not necessary to optimize the code in any way. When implemented on GPU, a network trained with Painting-by-Numbers is approx. 2.0 % slower than training without Painting-by-Numbers.

2 Cityscapes-C

Fig. 1 illustrates the utilized ImageNet-C[3] corruptions to generate Cityscapes-C. The severity level, i.e., the degree of the respective corruption, for each image is in this Figure mostly four or five. To illustrate the varying severity levels, we show in Fig. 2 the first three severity levels of a candidate of category *blur*, *noise*, *digital*, and *weather*.

[3] See https://github.com/hendrycks/robustness/tree/master/ImageNet-C for the implementation of ImageNet-C.

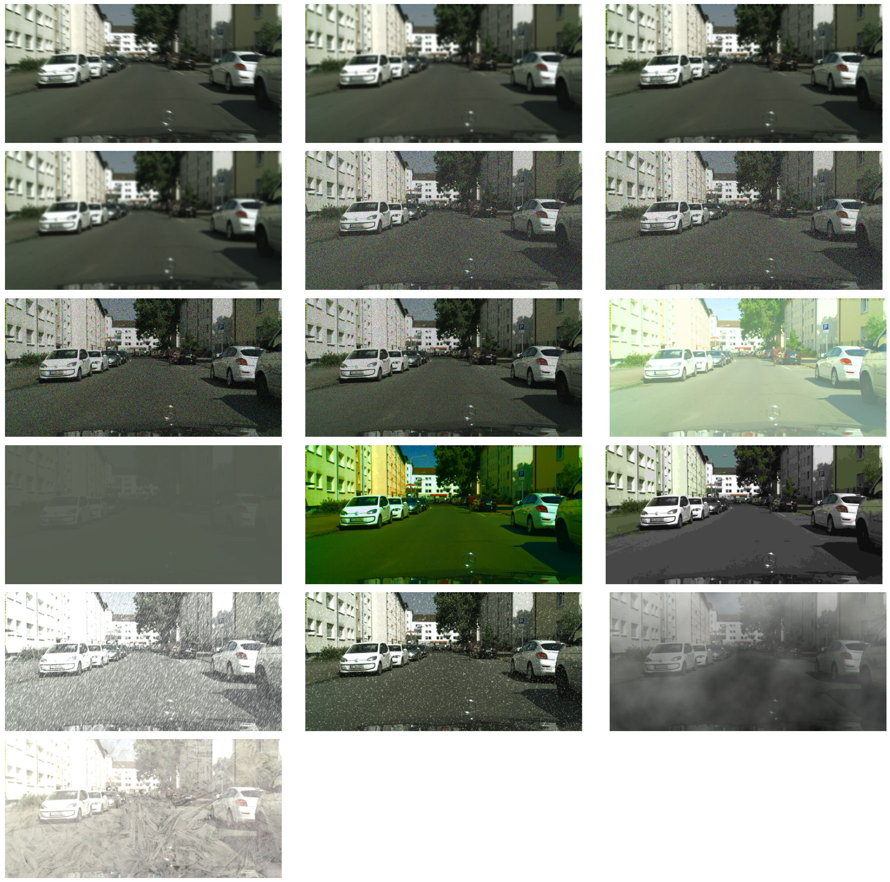

Fig. 1. Illustration of Cityscapes-C using many transformations of ImageNet-C. First row: Motion blur, defocus blur, frosted glass blur. Second row: Gaussian blur, Gaussian noise, impulse noise. Third row: Shot noise, speckle noise, brightness. Fourth row: Contrast, saturate, JPEG. Fifth row: Snow, spatter, fog. Sixth row: frost

3 Corruptions of Category *Digital* and *Weather*

The results in the main paper indicate that Painting-by-Numbers does not (compared to a regularly trained network) increase the performance with respect to the corruption *JPEG compression*. We explain this behavior due to the posterization effect of the JPEG compression algorithm (see Fig. 3). The Figure illustrates how the JPEG compression algorithm posterizes larger areas of the *car* and *road*. This is somewhat contrary to Painting-by-Numbers: Whereas our training schema (i.e., Painting-by-Numbers) alpha-blends the image with a homogeneous, texture-free representation of a class, the JPEG compression causes new, non-distinct shapes within a class instance, see Fig. 3 b.

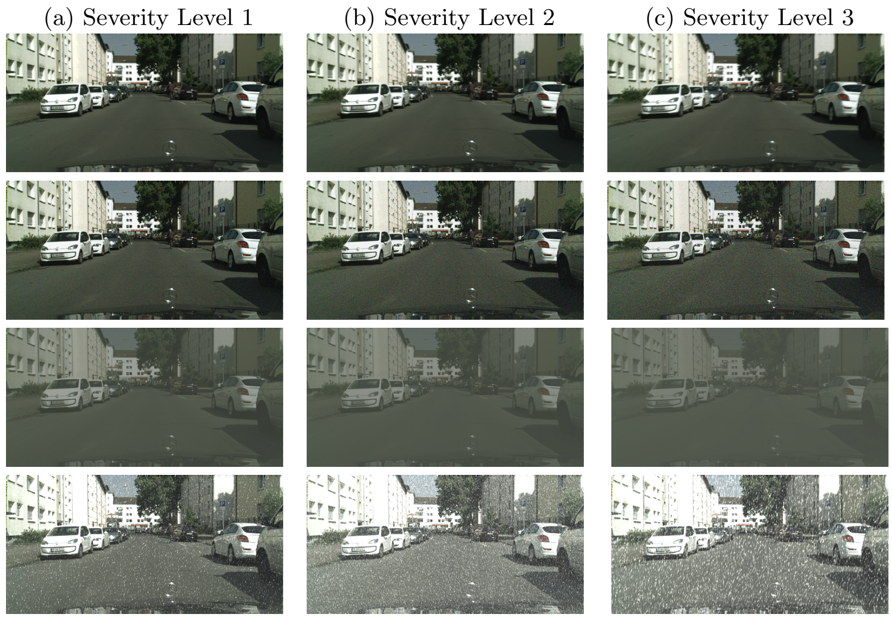

Fig. 2. Illustration of the first three severity levels of Cityscapes-C for a candidate of the categories *blur*, *noise*, *digital*, and *weather*. First row: Motion blur. Second row: Gaussian noise. Third row: Contrast. Fourth row: Snow

Though our model's higher performance with respect to image corruptions of category weather is less than, for example, for image noise, the predictions of a network trained with Painting-by-Numbers are improved for key-classes of category "things" such as *cars*, and *persons*. Table 1 lists the individual IoU score for both the reference model (i.e., the network was trained on clean data only) and our model (i.e., the network was trained with Painting-by-Numbers). We used for both models the ResNet-50 network backbone.

For example, with respect to *spatter*, the IoU score for classes *car*, and *person* is significantly higher by 46.6 %, and 31.3 %. Whereas our model struggles with respect to several classes for corruption *snow*, its IoU for "things" as *person*, and *rider* is higher by more than 7.0 %. For image corruption *fog*, our model performs better for almost every class. With respect to frost, the IoU score for *cars* of our model is higher by 8.3 %.

4 Detailed Quantitative Results

Each image corruption of Cityscapes-C is parameterized in five severity levels. Table 1 from the main paper contains the average score for these five severity

(a) JPEG compressed validation image (b) Zoom of (a)

Fig. 3. A validation image of Cityscapes corrupted by *JPEG compression* and a respective zoom in (b). The crop in (b) visualizes the posterization effect of *JPEG compression*. Whereas Painting-by-Numbers alpha-blends the image with a homogeneous, texture-free representation of a class, the JPEG compression causes new, non-distinct shapes within a class

Table 1. IoU for each class of several candidates of category *weather* evaluated on the Cityscapes dataset. The IoU score of *spatter* of our model for classes *car* and *person* is significantly higher (46.6 % and 31.3 %, respectively) than the IoU of the reference model. Overall, we see many bold numbers for "things" of our Painting-by-Numbers model

	road	sidewalk	building	wall	fence	pole	traffic light	traffic sign	vegetation	terrain	sky	person	rider	car	truck	bus	train	motorcycle	bicycle	mean
Reference (ResNet-50)																				
Snow	**81.2**	**16.6**	**60.0**	1.4	**3.2**	**21.2**	**15.9**	**34.3**	**40.4**	**20.6**	**70.0**	25.6	5.7	24.4	8.4	12.0	3.6	**3.5**	**35.6**	**25.5**
Spatter	48.5	6.5	55.2	**7.5**	12.5	30.7	**35.1**	30.3	**66.1**	16.3	3.9	19.4	15.1	26.7	1.8	17.9	1.2	14.6	34.6	23.4
Fog	93.2	60.5	79.2	18.6	35.3	**46.0**	40.4	**63.9**	73.8	7.8	77.9	69.2	46.7	85.6	52.3	68.8	47.1	45.3	66.4	56.7
Frost	**54.7**	**12.7**	38.7	1.0	**13.1**	13.6	11.0	38.5	41.8	**12.2**	40.8	21.1	8.1	32.3	7.0	8.6	3.9	0.1	23.6	20.1
Our (ResNet-50)																				
Snow	59.8	4.3	47.5	**1.5**	2.8	10.5	11.5	29.7	26.3	9.8	67.3	**32.7**	**8.4**	**32.1**	**12.0**	**15.7**	**7.3**	1.6	30.1	21.6
Spatter	**79.9**	**24.9**	**69.3**	5.9	**27.1**	**36.8**	33.7	**39.4**	61.3	**24.7**	**42.6**	**50.7**	**23.1**	**73.3**	**20.2**	**30.1**	**7.2**	**19.6**	**50.0**	**37.9**
Fog	**95.7**	**70.5**	**84.8**	**32.3**	**46.6**	44.0	**47.4**	62.9	**84.8**	**33.0**	**87.8**	**70.5**	**50.7**	**90.6**	**62.1**	**79.9**	**68.6**	**51.5**	**67.6**	**64.8**
Frost	51.6	10.9	**49.8**	**2.0**	11.8	**15.4**	**15.4**	**42.5**	**50.9**	8.2	**58.3**	**24.2**	**13.7**	**40.6**	**8.1**	**15.7**	**9.3**	**1.8**	**35.4**	**24.5**

levels[4]. For completeness, we provide in this section the mIoU score for each severity level. Please find the mIoU scores for severity level 1−5 in Table 2 − Table 6, respectively.

5 Understanding Painting-by-Numbers

In section 4.3 in the main paper, we conducted several experiments to verify an increased shape bias of a network which is trained with Painting-by-Numbers. In one of these experiments, we removed the texture of a class and replaced it by the data-wide RGB-mean of the respective class. We provide supplementary results of an extended version of this experiment, where we further black the

[4] Except for image noise, where we only considered severity levels that the Signal-to-Noise ratio is larger than 10

Table 2. Results on the Cityscapes dataset. Each entry shows the mean IoU of several corrupted variants of the Cityscapes dataset for **severity level 1**. The higher mIoU of either the reference model or the respective model trained with Painting-by-Numbers is bold. Overall, we see many more bold numbers for our Painting-by-Numbers model

	Blur				**Noise**				**Digital**				**Weather**			
Network	Motion	Defocus	Frosted Glass	Gaussian	Gaussian	Impulse	Shot	Speckle	Brightness	Contrast	Saturate	JPEG	Snow	Spatter	Fog	Frost
Reference																
MobileNet-V2	**69.0**	**68.4**	**65.8**	**71.6**	9.6	14.2	11.1	32.1	68.4	66.5	54.8	**34.7**	22.5	**71.9**	56.0	36.1
ResNet-50	**72.8**	**72.4**	69.3	**75.1**	10.7	13.4	18.8	44.2	73.7	72.1	66.6	**44.5**	**25.5**	**76.7**	63.9	43.7
ResNet-101	71.6	70.6	69.3	73.1	22.9	22.9	30.9	50.8	72.9	71.9	67.4	**52.4**	24.3	75.9	64.1	46.9
Xception-41	73.6	72.5	70.2	75.2	27.6	25.1	38.1	57.5	75.7	73.2	71.3	**60.2**	**41.1**	**76.8**	66.9	45.9
Xception-71	74.2	72.8	70.6	75.4	22.0	11.5	32.3	54.4	76.3	73.9	71.2	**56.9**	36.7	76.7	70.2	46.6
Painting-by-Numbers																
MobileNet-V2	67.7	66.9	64.2	70.2	**17.4**	**18.4**	**21.3**	**40.7**	**70.0**	**68.6**	**62.8**	29.8	**25.9**	71.0	**66.4**	**42.9**
ResNet-50	72.3	72.3	**72.4**	74.3	**35.7**	**34.3**	**44.4**	**60.7**	**74.5**	**73.5**	**71.7**	38.3	21.6	75.4	**69.6**	**47.4**
ResNet-101	**73.1**	**72.7**	**71.8**	**74.8**	**36.5**	**39.6**	**46.2**	**63.1**	**75.7**	**74.3**	**72.8**	48.6	**24.5**	**76.0**	**71.0**	**50.4**
Xception-41	**74.2**	**73.4**	**72.1**	**75.2**	**46.2**	**39.9**	**53.1**	**64.8**	**76.8**	**74.6**	**73.6**	39.0	38.2	76.7	**71.9**	**46.9**
Xception-71	**75.6**	**75.3**	**73.6**	**76.7**	**35.5**	**38.4**	**45.3**	**61.7**	**78.2**	**76.2**	**76.2**	43.0	**41.6**	**78.2**	**74.6**	**51.8**

Table 3. Results on the Cityscapes dataset. Each entry shows the mean IoU of several corrupted variants of the Cityscapes dataset for **severity level 2**. The higher mIoU of either the reference model or the respective model trained with Painting-by-Numbers is bold. Overall, we see many more bold numbers for our Painting-by-Numbers model

	Blur				**Noise**				**Digital**				**Weather**			
Network	Motion	Defocus	Frosted Glass	Gaussian	Gaussian	Impulse	Shot	Speckle	Brightness	Contrast	Saturate	JPEG	Snow	Spatter	Fog	Frost
Reference																
MobileNet-V2	**64.9**	**64.9**	**58.7**	**67.1**	8.1	8.7	8.6	19.1	59.6	61.6	46.3	**23.4**	8.4	**57.5**	51.1	16.5
ResNet-50	68.7	69.6	61.0	71.4	3.3	3.4	5.4	31.2	67.3	68.5	62.9	**29.1**	**11.1**	54.8	60.1	20.1
ResNet-101	68.0	68.2	62.4	69.9	10.6	10.5	15.3	40.3	66.8	69.2	63.7	**39.0**	9.5	**58.3**	59.7	25.2
Xception-41	71.0	70.1	64.6	71.5	12.6	11.4	18.1	49.2	71.7	70.4	68.2	**48.5**	**18.0**	60.6	62.4	**24.0**
Xception-71	72.3	70.4	65.8	72.0	9.2	7.0	10.9	43.0	73.3	71.3	67.6	**43.5**	12.8	**63.6**	67.3	22.7
Painting-by-Numbers																
MobileNet-V2	63.7	62.6	57.4	65.3	**11.4**	**11.1**	**12.2**	**30.8**	**66.3**	**66.0**	**57.9**	21.3	**8.9**	54.2	**63.9**	**23.5**
ResNet-50	**70.0**	**70.1**	**67.9**	**71.8**	**23.1**	**21.5**	**27.7**	**52.6**	**71.8**	**71.8**	**69.2**	26.0	7.3	**58.0**	**67.7**	**24.5**
ResNet-101	**70.8**	**71.1**	**65.6**	**72.5**	**19.7**	**23.2**	**26.5**	**54.3**	**73.3**	**72.9**	**71.3**	35.5	**9.9**	56.0	**68.6**	**28.3**
Xception-41	**72.7**	**70.6**	**65.8**	**72.7**	**30.0**	**25.1**	**35.3**	**58.3**	**75.3**	**72.5**	**69.0**	25.8	12.4	**63.7**	**69.6**	22.6
Xception-71	**74.1**	**73.8**	**68.2**	**75.3**	**17.6**	**18.4**	**23.1**	**53.6**	**77.0**	**74.7**	**72.7**	27.3	**16.1**	63.2	**72.5**	**28.2**

Table 4. Results on the Cityscapes dataset. Each entry shows the mean IoU of several corrupted variants of the Cityscapes dataset for **severity level 3**. The higher mIoU of either the reference model or the respective model trained with Painting-by-Numbers is bold. Overall, we see many more bold numbers for our Painting-by-Numbers model

	Blur				**Noise**				**Digital**				**Weather**			
Network	Motion	Defocus	Frosted Glass	Gaussian	Gaussian	Impulse	Shot	Speckle	Brightness	Contrast	Saturate	JPEG	Snow	Spatter	Fog	Frost
Reference																
MobileNet-V2	**55.4**	**50.4**	**39.7**	**54.9**	7.1	7.7	7.5	9.7	50.3	51.9	51.6	**18.8**	9.8	**41.7**	46.0	10.4
ResNet-50	59.6	59.5	37.1	63.5	1.3	1.6	1.4	9.2	59.8	61.0	64.9	**22.4**	**10.7**	35.0	56.7	12.1
ResNet-101	62.4	60.9	41.3	64.1	5.0	6.7	8.5	18.9	58.8	63.9	62.3	**32.8**	9.9	40.6	55.3	16.5
Xception-41	64.5	**61.0**	**47.0**	**64.5**	4.6	6.8	6.6	22.7	65.8	65.5	70.6	**41.6**	**15.7**	42.0	58.5	**16.1**
Xception-71	65.5	**63.5**	**50.8**	65.6	4.3	5.6	5.1	14.5	69.7	65.9	73.3	**38.2**	13.4	42.2	63.4	13.1
Painting-by-Numbers																
MobileNet-V2	52.6	41.7	36.5	48.6	**8.6**	**9.1**	**9.1**	**15.7**	**62.8**	**59.9**	**66.8**	17.0	**10.8**	41.1	**60.7**	**17.3**
ResNet-50	**63.2**	**61.5**	**43.9**	**65.7**	**14.4**	**16.2**	**18.3**	**33.2**	**68.7**	**68.9**	**71.2**	20.8	8.7	**40.6**	**64.8**	**16.3**
ResNet-101	**64.0**	**64.4**	**45.7**	**67.5**	**8.7**	**15.1**	**11.9**	**29.9**	**70.6**	**69.6**	**74.4**	28.6	**10.7**	**40.9**	**65.8**	**20.1**
Xception-41	**67.1**	56.7	44.4	61.5	**16.5**	**17.6**	**21.2**	**39.3**	**73.2**	**67.6**	**76.3**	19.4	13.5	**48.1**	**66.1**	13.9
Xception-71	**67.9**	63.4	44.4	**67.1**	**8.8**	**10.8**	**11.1**	**29.0**	**75.3**	**71.2**	**77.6**	20.8	**14.3**	**48.3**	**69.7**	**19.3**

Table 5. Results on the Cityscapes dataset. Each entry shows the mean IoU of several corrupted variants of the Cityscapes dataset for **severity level 4**. The higher mIoU of either the reference model or the respective model trained with Painting-by-Numbers is bold. Overall, we see many more bold numbers for our Painting-by-Numbers model

	Blur				**Noise**				**Digital**				**Weather**			
Network	Motion	Defocus	Frosted Glass	Gaussian	Gaussian	Impulse	Shot	Speckle	Brightness	Contrast	Saturate	JPEG	Snow	Spatter	Fog	Frost
Reference																
MobileNet-V2	**39.6**	**31.9**	**33.1**	**34.9**	6.1	6.6	5.8	8.3	41.2	29.1	5.5	**14.0**	7.6	32.2	46.1	9.6
ResNet-50	45.3	**43.8**	31.7	49.1	1.2	1.2	1.2	3.7	53.0	41.7	8.7	**12.5**	**8.8**	23.4	54.6	10.8
ResNet-101	**49.4**	**45.5**	36.0	51.4	3.6	4.4	**5.3**	12.9	50.4	49.9	8.4	**22.4**	7.8	32.6	54.3	15.5
Xception-41	51.4	**42.1**	**42.0**	**44.8**	2.8	5.0	2.5	12.6	58.7	**50.7**	21.6	**27.0**	**10.9**	35.6	56.4	**14.9**
Xception-71	**53.5**	**50.6**	**44.5**	**53.8**	3.0	4.0	2.7	7.6	63.1	49.0	10.4	**25.3**	**9.8**	38.8	62.4	12.1
Painting-by-Numbers																
MobileNet-V2	35.1	22.6	27.6	23.8	**7.0**	**7.1**	**7.1**	**11.1**	**59.0**	**39.2**	**38.8**	10.8	**8.2**	**41.6**	**54.1**	**15.6**
ResNet-50	**46.5**	41.2	**36.4**	**50.6**	**8.9**	**9.7**	**9.8**	**25.7**	**66.0**	**60.0**	**49.4**	12.7	7.5	**37.9**	**59.5**	**14.8**
ResNet-101	49.2	44.7	**40.1**	**53.1**	**5.1**	**6.3**	5.1	**19.2**	**67.7**	**62.3**	**58.9**	17.6	**8.7**	**36.4**	**59.5**	**17.8**
Xception-41	**51.5**	31.4	36.4	27.7	**9.7**	**9.7**	**11.0**	**29.5**	**70.9**	50.0	**63.8**	10.7	8.9	**49.1**	**59.4**	11.9
Xception-71	52.8	39.4	35.5	37.7	**5.1**	**5.3**	**5.9**	**18.4**	**73.4**	**57.1**	**64.4**	12.1	9.6	**57.3**	**62.9**	**16.3**

Table 6. Results on the Cityscapes dataset. Each entry shows the mean IoU of several corrupted variants of the Cityscapes dataset for **severity level 5**. The higher mIoU of either the reference model or the respective model trained with Painting-by-Numbers is bold

	Blur				**Noise**				**Digital**				**Weather**			
Network	Motion	Defocus	Frosted Glass	Gaussian	Gaussian	Impulse	Shot	Speckle	Brightness	Contrast	Saturate	JPEG	Snow	Spatter	Fog	Frost
Reference																
MobileNet-V2	**33.3**	**19.2**	**26.4**	**12.2**	5.2	5.7	5.2	7.0	32.3	10.0	4.2	**10.8**	5.7	13.5	**39.4**	7.8
ResNet-50	**39.1**	**31.0**	27.4	**23.2**	1.4	1.3	1.3	1.4	45.3	20.3	5.4	**8.7**	**8.4**	9.1	**45.8**	8.3
ResNet-101	**43.2**	**31.0**	30.1	**23.2**	**3.8**	**4.7**	**4.8**	9.6	39.8	29.0	6.0	**16.0**	7.7	**20.1**	**45.3**	12.0
Xception-41	**46.3**	**26.6**	**33.5**	**12.9**	1.8	3.6	2.7	6.5	48.5	**30.0**	11.4	**17.8**	**8.4**	18.3	**47.9**	**11.7**
Xception-71	**47.1**	**35.4**	**31.3**	**21.8**	2.7	**3.4**	3.1	4.6	52.4	26.0	6.0	**16.8**	7.7	18.6	56.1	8.2
Painting-by-Numbers																
MobileNet-V2	28.2	13.3	17.9	7.2	**5.3**	**5.8**	**6.3**	**8.4**	**54.5**	**20.3**	**28.6**	9.0	**6.9**	**26.8**	37.7	**12.5**
ResNet-50	38.5	22.2	**30.8**	12.9	**5.8**	**6.5**	**6.5**	**19.5**	**63.0**	**46.9**	**41.3**	8.9	8.1	**18.6**	43.6	**11.3**
ResNet-101	40.8	27.5	**34.5**	17.2	3.7	4.0	3.9	**10.4**	**63.5**	**50.5**	**52.1**	12.2	**8.2**	19.4	43.3	**14.2**
Xception-41	44.4	14.2	25.4	3.5	**6.0**	**6.2**	**7.7**	**20.9**	**67.5**	29.9	**54.5**	7.0	7.9	**28.7**	41.6	9.2
Xception-71	45.0	16.1	21.1	4.3	**3.3**	3.3	**4.1**	**11.5**	**70.7**	**40.4**	**54.5**	7.9	**9.4**	**40.0**	**47.2**	**12.0**

remaining classes, as illustrated in Fig. 4. The sensitivity score for every class is listed in Table 7. As the results in the main paper indicate, the reference model is not able to segment any classes except *building* and *sky*, since it may rely mostly on the mean color of these classes. Our model, on the other hand, achieves considerably higher sensitivity scores for category "things" such as *road*, *traffic light*, *traffic sign*, *pole*, *person*, and *car*.

(a) Original image

(b) Silhouette of *car*

Fig. 4. An original image of the Cityscapes validation image (a) and the corresponding silhouette for class *car*. Similar to the main paper, the class texture is replaced by RGB-mean, but in addition, we also black the remaining classes. A network segmenting such an image is not able to rely on context or background information but has to utilize shape-based cues for the segmentation

Table 7. Sensitivity score per class for several corrupted variants on class-level of the Cityscapes dataset, using ResNet-50 as the network backbone. In these experiments, solely the silhouette of a class is present, forcing the network to rely mostly on the class-shape for the segmentation. Whereas the reference model is, in most cases, not able to segment the images, our model performs superior for many classes of category "things" with a distinct shape. The higher sensitivity score of a network backbone of either the reference (top) or our model (bottom) is bold. Overall, we see many more bold numbers for our Painting-by-Numbers model

	road	sidewalk	building	wall	fence	pole	traffic light	traffic sign	vegetation	terrain	sky	person	rider	car	truck	bus	train	motorcycle	bicycle
Reference (ResNet-50)	2.1	0.0	**93.3**	0.4	0.0	1.0	0.0	0.1	0.0	0.0	**67.4**	0.8	0.0	0.0	0.1	**5.6**	**2.2**	0.0	0.0
Our (ResNet-50)	**99.1**	**58.4**	18.1	**5.9**	**2.1**	**86.2**	**31.8**	**26.6**	**4.1**	**8.2**	61.9	**79.5**	**1.9**	**77.4**	**0.1**	0.0	0.0	**3.8**	**12.1**