URIE: Universal Image Enhancement for Visual Recognition in the Wild

— Supplementary Material —

Taeyoung Son[1] Juwon Kang[1] Namyup Kim[1] Sunghyun Cho[2] Suha Kwak[2]

[1]Department of Computer Science and Engineering
[2]Graduate School of Artificial Intelligence
POSTECH, Pohang, Korea
{ty.son, gjw0917, namyup, s.cho, suha.kwak}@postech.ac.kr

This supplementary material presents experimental results omitted from the main paper due to the space limit. Sec. 1 empirically justifies the combination of Batch Normalization (BN) [2] and Instance Normalization (IN) [4] in our Selective Enhancement Module (SEM). In Sec. 2, we verify the impact of the large-scale learning in terms of performance and universality of URIE by examining the same models trained with smaller datasets. Also, Sec. 3 explains why it is not straightforward to compare URIE with existing image restoration models in our setting for robust visual recognition, and Sec. 4 analyzes image restoration performance of URIE. Finally, Sec. 5 presents more qualitative results of object detection and semantic segmentation on the distorted PASCAL VOC datasets.

1 Ablation Study Regarding Different Normalizations

The two enhancement steps in SEM have different normalization operations, IN and BN. This section demonstrates that the combination of IN and BN comes with benefits due to their complementary roles in image enhancement. To this end, we design two variants of URIE, *URIE-BN* and *URIE-IN*, and compare them with the original one. Specifically, in URIE-BN all normalization operations are BN, and in URIE-IN those are all implemented as IN. For a fair comparison, the two models are trained in the same setting with URIE.

The three models are evaluated in terms of recognition performance on distorted images by following the protocol in the main paper. Tab. 1, 2, 3, and 4 summarize their performance on the four different recognition tasks. These results show that using only one of the normalization operations degrades the final recognition performance noticeably in almost all settings. To investigate the effect of IN and BN in more detail, we measure the performance of the three models per distortion type on the distorted CUB dataset. As shown in Fig. 1, URIE-BN and URIE-IN exhibit clearly different tendencies. URIE-IN works better than URIE-BN for noise-type distortions while URIE-BN dominates URIE-IN when images are corrupted by blur-type distortions or adverse weathers. On the other hand, URIE adopting both of IN and BN performs best for most of the distortion types. These results justify our assumption that the two different normalization operations are complementary to each other.

Table 1. Classification accuracy on the ImageNet dataset. The numbers in parentheses indicate the differences from the performance of URIE. V16, R50, and R101 denote VGG-16, ResNet-50, and ResNet-101, respectively.

	URIE			URIE-BN			URIE-IN		
	Clean	Seen	Unseen	Clean	Seen	Unseen	Clean	Seen	Unseen
V16	67.1	42.4	44.8	63.0 (-4.1)	41.3 (-1.1)	43.5 (-1.3)	63.9 (-3.2)	39.9 (-2.5)	44.4 (-0.4)
R50	72.9	55.1	56.5	70.4 (-2.5)	54.0 (-1.1)	55.0 (-1.5)	71.0 (-1.9)	52.8 (-2.3)	55.6 (-0.9)
R101	74.1	57.8	59.4	71.7 (-2.4)	56.8 (-1.0)	58.1 (-1.3)	72.1 (-2.0)	55.5 (-2.3)	58.1 (-1.3)

Table 2. Classification accuracy on the CUB dataset. The numbers in parentheses indicate the differences from the performance of URIE. V16, R50, and R101 denote VGG-16, ResNet-50, and ResNet-101, respectively.

	URIE			URIE-BN			URIE-IN		
	Clean	Seen	Unseen	Clean	Seen	Unseen	Clean	Seen	Unseen
V16	77.9	58.8	48.9	76.7 (-1.3)	56.2 (-2.2)	47.4 (-1.5)	77.0 (-0.9)	56.8 (-2.0)	49.8 (+0.9)
R50	83.7	64.7	54.9	82.4 (-1.3)	62.9 (-1.8)	54.3 (-0.6)	82.9 (-0.8)	62.1 (-2.6)	56.4 (+1.5)
R101	84.1	67.1	56.6	82.9 (-1.2)	65.0 (-2.1)	57.0 (+0.4)	82.6 (-1.5)	64.8 (-2.3)	57.4 (+0.8)

Table 3. Object detection performance of SSD 300 in mAP (%) on the VOC 2007 dataset. The numbers in parentheses indicate the differences from the scores of URIE.

URIE			URIE-BN			URIE-IN		
Clean	Seen	Unseen	Clean	Seen	Unseen	Clean	Seen	Unseen
76.5	59.4	62.7	74.4 (-2.1)	57.9 (-1.5)	61.0 (-1.7)	74.1 (-2.4)	56.9 (-2.5)	60.9 (-1.8)

Table 4. Semantic segmentation performance of DeepLab v3 in mIoU (%) on the VOC 2012 dataset. The numbers in parentheses indicate the differences from the score of URIE.

URIE			URIE-BN			URIE-IN		
Clean	Seen	Unseen	Clean	Seen	Unseen	Clean	Seen	Unseen
78.6	67.0	67.2	78.4 (-0.2)	66.3 (-0.7)	68.0 (+0.8)	78.5 (-0.1)	64.9 (-2.1)	67.4 (+0.2)

2 Impact of Large Scale Training

To demonstrate the advantage of large-scale training using the ImageNet dataset, we compare URIE with its other two variants, *URIE-1/4* and *URIE-1/16*, trained using a quarter and a sixteenth of the corrupted ImageNet dataset, respectively. Tab. 6, 7, 8, and 9 summarize the performance of the three models, and show that URIE-1/4 and URIE-1/16 degrade the recognition performance substantially and are not well transferred to other tasks compared to the original URIE. These results justify our assumption that training on a large-scale distorted image dataset can improve the performance and universality of URIE.

3 Practical Issues on Comparisons to Restoration Models

To prove effectiveness of proposed method, it would be best to compare URIE to image restoration models trained in the same manner with URIE. However, we would stress that it is often impractical to train and evaluate them in the same

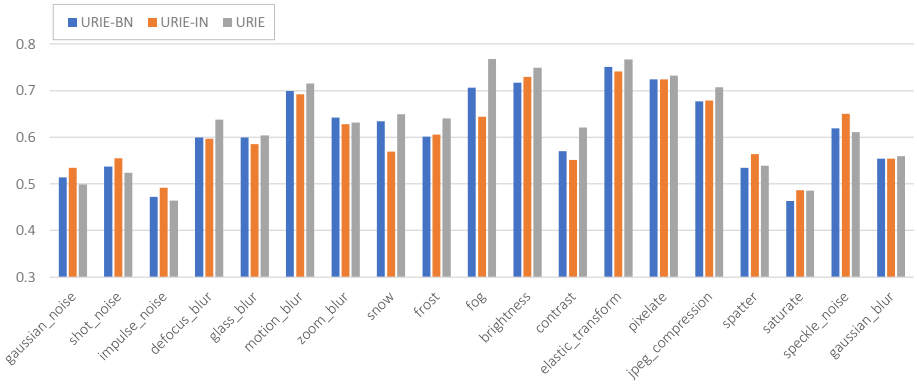

Fig. 1. Performance comparison between URIE-BN, URIE-IN, and URIE per distortion type on the corrupted CUB dataset.

Table 5. Restoration performance on the CUB dataset in terms of MSE and SSIM. Their accuracies over CUB dataset are presented alongside.

	MSE	SSIM	Recog. Acc.
URIE	0.331	0.358	55.1
URIE-MSE	0.158	0.445	44.5
URIE-SSIM	0.206	0.445	44.6
OWAN [3]	0.380	0.366	42.6

setting. In our setting, specifically, URIE is trained with the recognition-aware loss on the ImageNet-C dataset whose images are corrupted by diverse and latent distortions. Hence models must be (1) efficient in computation and memory usage and (2) able to deal with a multitude of latent distortion types. Unfortunately, most prior studies on image restoration do not meet the two conditions since they rely on considerably heavier networks and assume a single distortion type. In particular, we found that it takes impractically long time to train such enhancement networks with the recognition-aware loss on the ImageNet-C dataset (*e.g.*, taking 137 days on 4 Tesla P40 GPUs in the case of OWAN [3]).

4 Restoration Performance of URIE and Its Variants

This section presents restoration performance of URIE and its variants. They are evaluated in terms Mean Squared Error (MSE) and Structural SIMilarity (SSIM). In addition to three models used in our experiment, we consider another variant of URIE that is trained with SSIM loss, called URIE-SSIM. As reported in the Tab. 5, URIE is worse than URIE-MSE and URIE-SSIM in restoration but substantially outperforms them in recognition, which suggests that URIE works as desired. Also, URIE and its variants outperform OWAN [3] in recognition performance. This is partly due to the superiority of our network architecture,

Table 6. Classification accuracy on the ImageNet dataset. The numbers in parentheses indicate the differences from the performance of URIE. V16, R50, and R101 denote VGG-16, ResNet-50, and ResNet-101, respectively.

	URIE			URIE-1/4			URIE-1/16		
	Clean	Seen	Unseen	Clean	Seen	Unseen	Clean	Seen	Unseen
V16	67.1	42.4	44.8	64.9 (-2.2)	39.3 (-5.0)	43.2 (-1.6)	64.3 (-2.8)	35.1 (-7.3)	40.6 (-4.2)
R50	72.9	55.1	56.5	71.5 (-1.4)	51.5 (-2.5)	54.1 (-1.0)	71.0 (-1.9)	47.3 (-7.8)	51.6 (-4.9)
R101	74.1	57.8	59.4	72.8 (-1.3)	54.9 (-1.9)	57.2 (-2.2)	72.4 (-1.7)	51.2 (-6.6)	55.8 (-3.6)

Table 7. Classification accuracy on the CUB dataset. The numbers in parentheses indicate the differences from the performance of URIE. V16, R50, and R101 denote VGG-16, ResNet-50, and ResNet-101, respectively.

	URIE			URIE-1/4			URIE-1/16		
	Clean	Seen	Unseen	Clean	Seen	Unseen	Clean	Seen	Unseen
V16	77.9	58.8	48.9	76.8 (-1.1)	57.1 (-1.7)	48.0 (-0.9)	76.8 (-1.1)	53.7 (-5.1)	45.8 (-3.1)
R50	83.7	64.7	54.9	83.2 (-0.5)	62.4 (-2.3)	54.5 (-0.4)	82.7 (-1.0)	58.4 (-6.3)	53.4 (-1.5)
R101	84.1	67.1	56.6	83.2 (-0.9)	64.8 (-2.3)	56.3 (-0.3)	83.5 (-0.6)	61.7 (-5.4)	54.5 (-2.1)

Table 8. Object detection performance of SSD 300 in mAP (%) on the VOC 2007 dataset. The numbers in parentheses indicate the differences from the scores of URIE.

URIE-BN			URIE-1/4			URIE-1/16		
Clean	Seen	Unseen	Clean	Seen	Unseen	Clean	Seen	Unseen
76.5	59.4	62.7	75.3 (-1.2)	56.8 (-2.6)	61.1 (-1.6)	75.2 (-1.3)	53.0 (-6.4)	59.4 (-3.3)

Table 9. Semantic segmentation performance of DeepLab v3 in mIoU (%) on the VOC 2012 dataset. The numbers in parentheses indicate the differences from the score of URIE.

URIE			URIE-1/4			URIE-1/16		
Clean	Seen	Unseen	Clean	Seen	Unseen	Clean	Seen	Unseen
78.6	67.0	67.2	78.5 (-0.1)	65.1 (-1.9)	67.3 (+0.1)	78.4 (-0.2)	62.7 (-4.3)	65.3 (-1.9)

which better handles diverse distortions and images captured in uncontrolled environments.

5 More Qualitative Results

This section presents more qualitative results omitted in the main paper due to the space limit. Fig. 2 and Fig. 3 show the results of URIE on the PASCAL VOC 2012 [1] semantic segmentation dataset. They show that URIE can enhance images while focusing more on object-like areas instead of background. Also, the results not always looking plausible, especially when compared to those of the other methods, but directly improve the performance of the following semantic segmentation model. Fig. 4 and Fig. 5 exhibit results on the PASCAL VOC 2007 [1] object detection dataset. Likewise, URIE best recovers salient regions of the images and improves the robustness of the object detector.

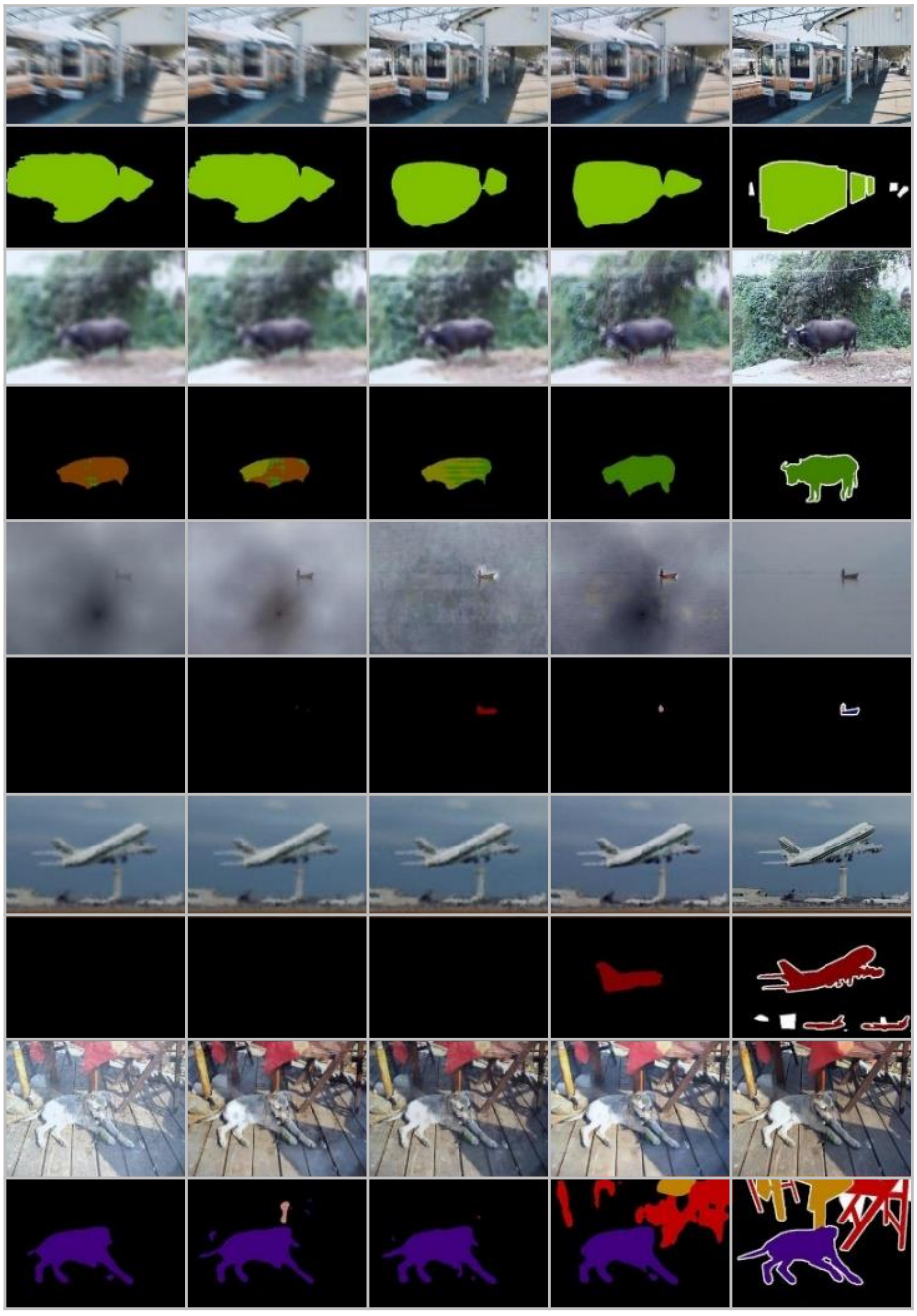

Fig. 2. Additional qualitative results of DeepLab v3 on the VOC 2012 dataset. (a) Corrupted input. (b) OWAN [3]. (c) URIE-MSE. (d) URIE. (e) Ground-truth.

Fig. 3. Additional qualitative results of DeepLab v3 on the VOC 2012 dataset. (a) Corrupted input. (b) OWAN [3]. (c) URIE-MSE. (d) URIE. (e) Ground-truth.

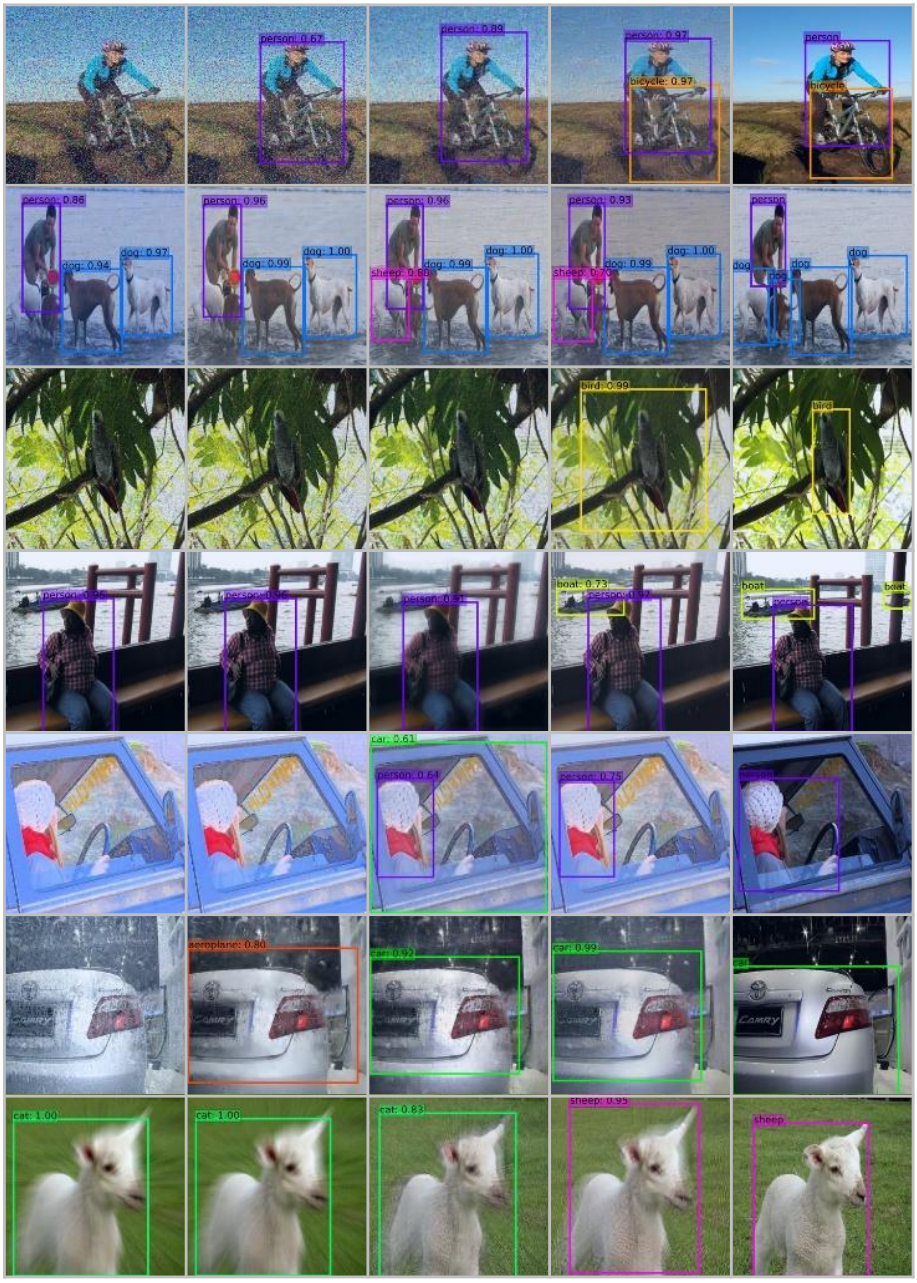

Fig. 4. Additional qualitative results of SSD 300 on the VOC 2007 dataset. (a) Corrupted input. (b) OWAN [3]. (c) URIE-MSE. (d) URIE. (e) Ground-truth.

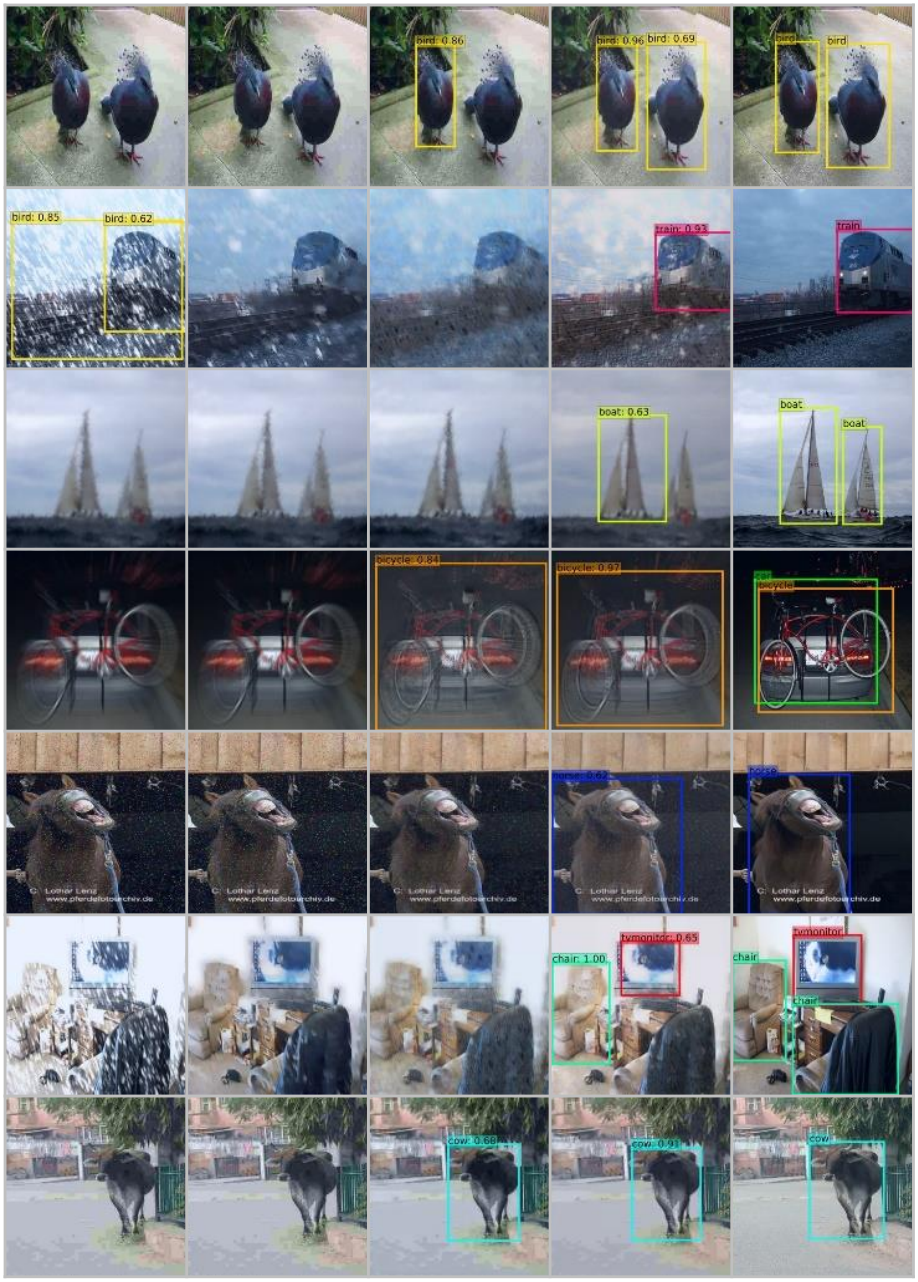

Fig. 5. Additional qualitative results of SSD 300 on the VOC 2007 dataset. (a) Corrupted input. (b) OWAN [3]. (c) URIE-MSE. (d) URIE. (e) Ground-truth.

References

1. Mark Everingham, Luc Van Gool, Christopher KI Williams, John Winn, and Andrew Zisserman. The Pascal Visual Object Classes (VOC) Challenge. *International Journal of Computer Vision (IJCV)*, 2010. 4
2. Sergey Ioffe and Christian Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *Proc. International Conference on Machine Learning (ICML)*, 2015. 1
3. Masanori Suganuma, Xing Liu, and Takayuki Okatani. Attention-based adaptive selection of operations for image restoration in the presence of unknown combined distortions. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019. 3, 5, 6, 7, 8
4. Dmitry Ulyanov, Andrea Vedaldi, and Victor Lempitsky. Instance normalization: The missing ingredient for fast stylization. *arXiv preprint arXiv:1607.08022*, 2016. 1