2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)

SimUSR: A Simple but Strong Baseline for Unsupervised Image Super-resolution

Namhyuk Ahn[†*]
Ajou University
aa0dfg@ajou.ac.kr

Jaejun Yoo[*]
EPFL
jaejun.yoo88@gmail.com

Kyung-Ah Sohn[‡]
Ajou University
kasohn@ajou.ac.kr

Abstract

In this paper, we tackle a fully unsupervised super-resolution problem, i.e., neither paired images nor ground truth HR images. We assume that low resolution (LR) images are relatively easy to collect compared to high resolution (HR) images. By allowing multiple LR images, we build a set of pseudo pairs by denoising and downsampling LR images and cast the original unsupervised problem into a supervised learning problem but in one level lower. Though this line of study is easy to think of and thus should have been investigated prior to any complicated unsupervised methods, surprisingly, there are currently none. Even more, we show that this simple method outperforms the state-of-the-art unsupervised method with a dramatically shorter latency at runtime, and significantly reduces the gap to the HR supervised models. We submitted our method in NTIRE 2020 super-resolution challenge and won 1st in PSNR, 2nd in SSIM, and 13th in LPIPS. This simple method should be used as the baseline to beat in the future, especially when multiple LR images are allowed during the training phase. However, even in the zero-shot condition, we argue that this method can serve as a useful baseline to see the gap between supervised and unsupervised frameworks.

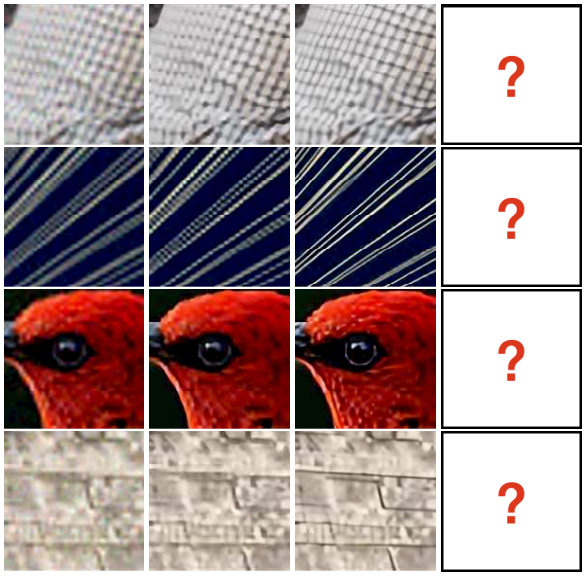

Figure 1. NTIRE 2020 Super-resolution challenge (×4) [18] images (1st column). ZSSR [21] with BM3D [4] results (2nd column). Our SimUSR with BM3D results (3rd column). Ground truth HR image (4th column) is not available in this setup. Our method achieves superior SR performance for all the cases.

1. Introduction

Single image super-resolution (SISR) is a longstanding task in computer vision area, which focuses on recovering a high-resolution (HR) image from a single low-resolution (LR) image. Since this task has to solve a one-to-many mapping problem, building an effective SISR method is challenging. Despite the difficulties, thanks to the developments of deep learning and large-scale datasets with high quality images, many learning-based SR methods [2, 5, 12, 14, 15, 29] have recently shown prominent performance gains over traditional optimization-based approaches [23]. With the power of the deep neural network, they enjoy the dramatic boost of the SR performance by stacking a lot of layers [29] or widen the network [15].

While most of the deep learning-based SR methods heavily rely on a large number of image pairs (supervised SR), unfortunately, such a large-scale and high quality dataset is not always accessible, especially when we deal with a real environment. A few recent works have proposed a workaround solving an unpaired SR task [7, 16, 27]. Since this setup does not require full supervision of LR and HR image pairs but images from each domain, it is a more realistic scenario for many real-world applications. However, in many applications (*e.g.*, medical image), gathering HR (or clean) images itself requires a lot of efforts or sometimes even impossible.

[†] Most work was done while in NAVER Corp.
[*] indicates equal contribution. ‡ indicates corresponding author.

2160-7516/20/$31.00 ©2020 IEEE
DOI 10.1109/CVPRW50498.2020.00245

Authorized licensed use limited to: EPFL LAUSANNE. Downloaded on October 02,2020 at 15:11:59 UTC from IEEE Xplore. Restrictions apply.

To address this, Shocher *et al*. [21] have proposed a fully unsupervised method, called zero-shot super-resolution (ZSSR), which performs both training and testing at runtime using only a single LR test image. By learning a mapping from a scale-down version of the LR image to itself, ZSSR learns to exploit internal image statistics to super-resolve the given image. It outperformed the previous internal SR methods [11] in a huge margin with a high flexibility because the model can easily be adapted to any unknown degradation or downsample kernel.

However, ZSSR has several drawbacks. **1)** It requires an online optimization procedure at runtime. Since it needs at least 1K steps (both forward and backward propagation), the latency of the ZSSR is extremely high. **2)** It is difficult to benefit from a large capacity network. Because ZSSR has to perform online training on a single image, the model should be able to quickly adapt to the given image while avoiding the overfitting issue, which limits ZSSR to use a shallow network architecture. **3)** When noises are present in LR images, ZSSR shows deteriorated performance because the model can never learn to denoise, and even after adding a denoising module, it suffers from its restrictive framework. **4)** It does not utilize any prior information at all, which is an excessively restrictive constraint. While collecting LR-HR image pairs are difficult, acquiring LR images only is relatively easy and feasible in many real-world scenarios. Since the internal-based SR methods generally show worse SR performance than the external-based models, it is desirable to exploit every available prior information as long as it stays in the unsupervised regime.

To mitigate these limitations, we propose a simple baseline for unsupervised SR (SimUSR) that relaxes the ZSSR into a supervised setting. Instead of using a single image, our SimUSR make pseudo-pairs using multiple LR images. To correctly guide the model, we employ BM3D [4] to remove noises from the LR images when preparing the pairs. Though these are very simple corrections, they bring several benefits: our framework can now exploit every benefit of supervised learning. Thanks to this pseudo-supervision, ample prior information enables a model to reduce the performance gap between the unsupervised (only LR is available) and the supervised setting (HR is available). SimUSR can utilize recently developed network architectures and techniques that provide huge performance gains (Figure 1). In addition, since the online training is not necessary, SimUSR can significantly reduce its runtime latency as well. The differences of the supervised SR, ZSSR, and SimUSR are summarized in Figure 2.

We argue that our assumption is fairly practical while still remaining under the unsupervised learning setup; we only use LR images. Our approach is meaningful in that it investigates the blind spot of the field that should have been addressed but overlooked.

Our contributions are summarized as follows:

1. We propose a simple but strong baseline for unsupervised SR (SimUSR) (NTIRE Challenge: 1st in PSNR, 2nd in SSIM, and 13th in LPIPS).

2. By casting the unsupervised SR into the supervised SR, SimUSR provides stable offline learning with a dramatically decreased latency at runtime.

3. We provide a comprehensive analysis on the effect of using various SR networks and techniques. By taking the state-of-the-art techniques and SR networks as our backbone, our method shows further enhancements.

2. Related work

Supervised Super-resolution. Recently, deep learning-based super-resolution models [2, 5, 12, 15, 29] have shown a dramatic leap over the traditional algorithms [23]. Most of the successful deep SR approaches fall into the supervised setting, where a network is trained on an external dataset having low- and high-resolution pairs. As long as the size of the dataset and the network capacity are large enough, it is well known that the supervised approach provides a better chance to enhance the SR performance [15, 29]. However, it is also true that their performance and generalizability deteriorate dramatically when the dataset size is small and when there exists mismatch between training and testing environments [26, 6]. To mitigate this issue, recent approaches focus on *blind SR*, which assumes that there exist LR and HR pairs but with unknown degradation and downsample kernel [9]. Unlike the aboves, our proposed method can train a network even when there are no LR and HR pairs.

Unpaired Super-resolution. A few recent works have addressed an unpaired SR task [7, 16, 27] that does not assume a paired setting. Since this setup does not require a full supervision, it is a more realistic scenario for many real-world applications. Most of the methods employ generative adversarial framework [8] so that a generator learns to map HR images into their distorted LR version. Using this generated pairs, they train an SR network in a supervised setting. However, in practice, there are cases where HR images are not even available, which requires a fully unsupervised SR.

Unsupervised Super-resolution. Though the unpaired SR is sometimes considered as an unsupervised SR, we first clarify that unsupervised SR should strictly denote the task without any supervision neither paired images nor HR images. Under this definition, there are only a handful of studies [24, 10, 21] and zero-shot super-resolution (ZSSR) [21] falls into this. ZSSR uses *LR sons* that are downsampled images of the given LR test image (a.k.a *LR father*). Using this pseudo pairs, they train the model in a supervised manner but only exploiting the internal statistics of the given test

Authorized licensed use limited to: EPFL LAUSANNE. Downloaded on October 02,2020 at 15:11:59 UTC from IEEE Xplore. Restrictions apply.

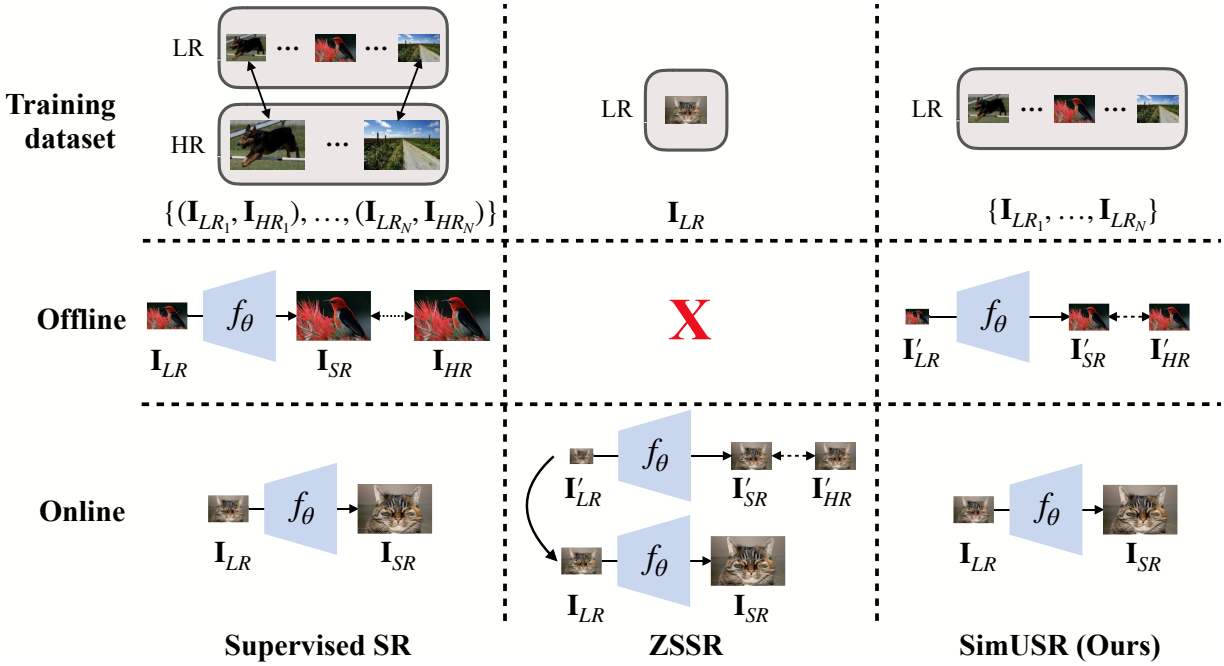

Figure 2. Schematic comparison of the supervised SR, ZSSR [21], and our SimUSR. We analyze current SR approaches in terms of the training dataset, offline phase, and online phase. The offline phase is operated beforehand the user's inference request (*i.e.* training process of the supervised SR). Online phase denotes runtime. (**1st row**) While supervised SR requires the LR-HR pairs, ZSSR and SimUSR use LR images only, making them more applicable to the real-world SR scenarios. ZSSR utilizes only a single test LR image and performs both optimization and inference at runtime. (**2nd, 3rd rows**) On the other hand, SimUSR exploits additional LR images and follows a similar procedure to the supervised setup, where the model is first trained offline and inference is done online.

image. Because every procedure is performed at runtime, ZSSR suffered from high latency. To overcome this, Soh *et al*. [22] have proposed meta-transfer ZSSR (MZSR). They added a meta-transfer learning phase to exploit the information of the external dataset, which decreased the number of the steps required at runtime. Still, to quickly optimize the network, MZSR was limited to use a simple 8-layer network. Unlike the aforementioned methods, our SimUSR can benefit from the larger capacities of recently developed SR models and short latency at runtime by removing the online update phase, while remaining in the fully unsupervised regime in that it only utilizes the LR images.

3. Zero-shot super-resolution

The zero-shot super-resolution (ZSSR) [21] tackles the fully unsupervised SR task, where only low-resolution images ($\mathbf{I}_{LR}$) are available. To do that, ZSSR performs both optimization and inference at runtime using a single test image (Figure 2). During the online optimizing phase, they use an test input image ($\mathbf{I}_{LR}$) as *LR father* ($\mathbf{I}_{LR}^{father}$) and generates *LR son* ($\mathbf{I}_{LR}^{son}$) by downsampling LR father with an arbitrary kernel k. By doing so, they create *pseudo-pair*

$$(\mathbf{I}'_{LR}, \mathbf{I}'_{HR}) = (\mathbf{I}_{LR}^{son}, \mathbf{I}_{LR}^{father}),$$

where $\mathbf{I}_{LR}^{son} = \mathbf{I}_{LR} \downarrow_{s,k}$ and $\mathbf{I}_{LR}^{father} = \mathbf{I}_{LR}$. Here, $\downarrow_{s,k}$ denotes a downsampling operation with an arbitrary kernel k and scale factor s.

With this pseudo-pair, optimizing a SR model now becomes a standard supervised setting. The core idea of ZSSR is to make the model learn internal image-specific statistics of a given test image during the online training. For inference, it generates final SR output ($\mathbf{I}_{SR}$) by feeding $\mathbf{I}_{LR}$ to the trained image-specific network.

4. Our method

We introduce a simple baseline for a fully unsupervised super-resolution task (SimUSR). Similar to the ZSSR [21], our method does not use any HR images for training the network. However, we slightly relax the constraint of ZSSR and assumes that it is relatively easy to collect the LR images, $\{\mathbf{I}_{LR_1}, \ldots, \mathbf{I}_{LR_N}\}$, where N is the number of LR images. This allows our method to exploit multiple pseudo-pairs:

$$(\mathbf{I}'_{LR_k}, \mathbf{I}'_{HR_k}) = (\mathbf{I}_{LR_k}^{son}, \mathbf{I}_{LR_k}^{father}), \quad \text{for } k = 1 \ldots N.$$

Here, we generate $\mathbf{I}_{LR}^{son}$ and $\mathbf{I}_{LR}^{father}$ with the same protocol that used in ZSSR.

Authorized licensed use limited to: EPFL LAUSANNE. Downloaded on October 02,2020 at 15:11:59 UTC from IEEE Xplore. Restrictions apply.

Table 1. Quantitative comparison (PSNR/SSIM) on the bicubic SR (scale $\times 4$) benchmark datasets. We boldface the best performance of both supervised SR and ours.

Dataset	Supervised SR			ZSSR	SimUSR (Ours)		
	CARN	RCAN	EDSR		CARN	RCAN	EDSR
Set5	32.13/0.8937	**32.63/0.9002**	32.46/0.8968	31.13/0.8796	31.94/0.8908	**32.40/0.8962**	32.37/0.8955
Set14	28.60/0.7806	**28.87/0.7889**	28.80/0.7876	28.01/0.7651	28.44/0.7786	**28.71/0.7860**	28.70/0.7855
B100	27.58/0.7349	**27.77/0.7436**	27.71/0.7420	27.12/0.7211	27.49/0.7324	**27.68/0.7394**	27.66/0.7389
Urban100	26.07/0.7837	**26.82/0.8087**	26.64/0.8033	24.61/0.7282	25.70/0.7740	**26.45/0.7986**	26.31/0.7940
Manga109	-	**31.22/0.9173**	31.02/0.9148	27.84/0.8657	30.03/0.9014	**30.73/0.9124**	30.59/0.9107

Table 2. Quantitative comparison (PSNR) on SR (scale $\times 4$) task with mixture of augmentation (MoA) [26]. We show the effect of MoA on our SimUSR and supervised SR (SSR) model. Note that SSR results are provided to show the improved upper limit again.

Type	Model	Set14	Urban	Manga
SimUSR	RCAN (+MoA)	28.80	26.60	30.85
SSR	RCAN (+MoA)	28.92	26.93	31.46

Though we now lose the generalizability over a single test image, compared to the cost of the relaxation, the benefits are very huge: we can fully enjoy the advantages of the supervised learning framework. More specifically, using these multiple pairs, we can now train a network offline and perform inference online as any supervised model usually does. Our method can be implemented by a simple modification of the supervised SR approach, it gives high flexibility and extensibility. For example, unlike the ZSSR and MZSR [22], which inevitably use shallow networks, we can use any off-the-shelf SR network and technique available, such as data augmentation [26] (Section 5.1). In addition, since the runtime of our SimUSR only depends on the network's inference speed, this also gives a huge acceleration in terms of the runtime latency (Section 5.3).

5. Experiments

In this section, we describe our experimental settings and compare the performance of our method with the supervised SR models and the ZSSR [21]. In Section 5.1, we analyze how much our SimUSR improves the performance over the ZSSR and how far we are left to reach the supervised performance. Then, in Section 5.2, we apply our method on the NTIRE 2020 SR dataset [18].

Baselines. We use ZSSR [21] as our major baseline method. However, since ZSSR and SimUSR are not designed to handle noisy cases, we attach BM3D [4] as a pre-processing step. For our SimUSR, we use various models as our backbone network. We use three SR models: CARN [2], RCAN [29] and EDSR [15]. Each of the model have different numbers of parameters from 1.1M to 43.2M (million).

Dataset and evaluation. We use the DF2K [1, 15] dataset for the bicubic degradation SR task. However, unlike the Lim *et al*. [15], we only use the LR images when we train the models. For evaluation, we use Set5 [3], Set14 [25], B100 [19], Urban100 [11], and Manga109 [20] for bicubic SR task. To evaluate our method on the real-world SR task, we use NTIRE 2020 dataset [18]. This dataset is generated with unknown degradation operation to simulate the realistic image processing artifacts. In addition, only non-paired LR and HR images are given so that the model should be trained via unsupervised setup. Same as DF2K, we do not use any of HR images at the training phase. We use PSNR and SSIM to measure the performance. We calculate both metrics on RGB channels for the NTIRE dataset while only using the Y channel for the bicubic SR task.

5.1. Bicubic super-resolution

Here, we compare SimUSR with the ZSSR and the supervised SR models. Though the classical bicubic SR task is not our main task, it provides a testbed to analyze every model simultaneously. This also shows how much gap there exists between the supervised and unsupervised frameworks. For fair comparison, we report our performance using different SR networks as our backbone (CARN [2], RCAN [29], and EDSR [15]). The quantitative comparison on various benchmark dataset is shown in Table 1. Exploiting the additional LR images, our SimUSR shows large improvements over the ZSSR in every case.

More interestingly, by exploiting the recent development of supervised SR techniques, such as data augmentation, SimUSR further reduces the gap toward the supervised learning models (Table 2). Note that, while the supervised models can use HR images as ground truth, SimUSR only uses LR images. Therefore, the model should generalize over the learned scale and pixel distributions. Toward this, we used mixture of augmentation (MoA) [26], which is a recent data augmentation method for low-level vision task. MoA is known to not only improve the performance but also enhance the generalization power of the model. By employing the MoA, which ZSSR does not benefit from (results not shown), our performance again increases by 0.09 dB (Set14), 0.15 dB (Urban100), and 0.12 dB (Manga109),

Authorized licensed use limited to: EPFL LAUSANNE. Downloaded on October 02,2020 at 15:11:59 UTC from IEEE Xplore. Restrictions apply.

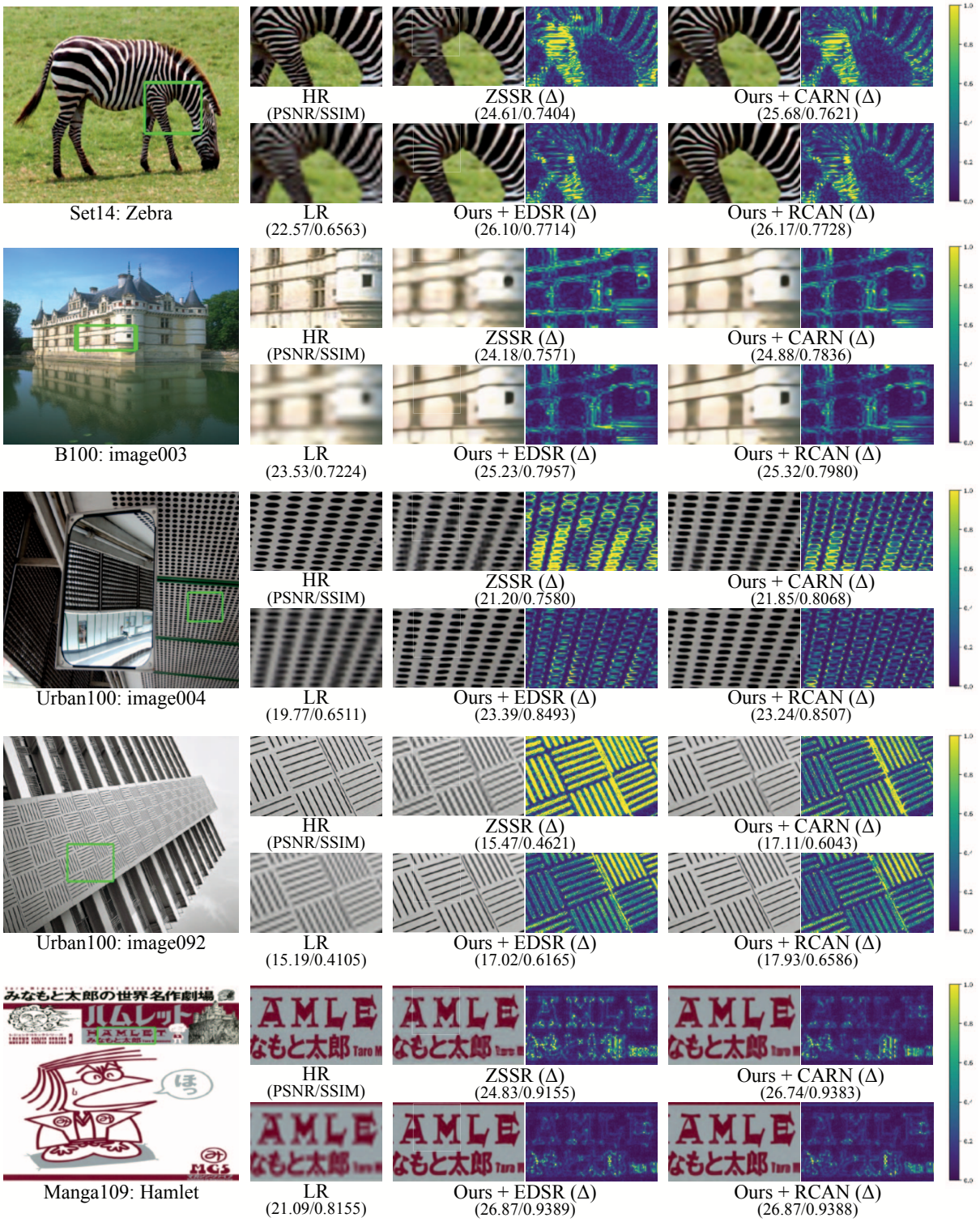

Figure 3. Qualitative comparison of using our proposed method on the various benchmark datasets which are generated by the Bicubic downsample kernel. Δ is the absolute residual intensity map between the network output and the ground-truth HR image.

Authorized licensed use limited to: EPFL LAUSANNE. Downloaded on October 02,2020 at 15:11:59 UTC from IEEE Xplore. Restrictions apply.

Table 3. Quantitative comparison (PSNR/SSIM) on the NTIRE 2020 dataset [18]. We analyze the effect of denoising (*w/ BM3D*) and affine transformations (*w/o Affine*). We also analyze the advantage of applying SimUSR.

Method	w/ BM3D	w/o Affine	PSNR / SSIM
ZSSR			25.82 / 0.6898
	✓		26.45 / 0.7320
	✓	✓	26.55 / 0.7344
SimUSR+CARN	✓	✓	27.19 / 0.7520
SimUSR+RCAN	✓	✓	27.24 / 0.7550
SimUSR+EDSR	✓	✓	**27.28 / 0.7554**

which are upto **3.63 dB** (Manga109) improvements over the ZSSR. Therefore, from now on, we use MoA with SimUSR by default unless it is specified.

The qualitative results also shows the superior results of SimUSR over the ZSSR (Figure 3). In all the cases, SimUSR benefits from the increased performance by using external LR images. This tendency is clearly shown in the residual intensity map between the SR and HR image. For example, our method (with any backbone) successfully restores the replicating patterns (1st, 3rd, and 4th rows) while ZSSR has difficulty of recovering distortions. Note that ZSSR is supposed to better learn the internal statistics by repeatedly seeing the same LR image patches, which is in principle good at recovering replicating patterns.

5.2. Real-world super-resolution

In this section, we compare ZSSR [21] and our method on the NTIRE 2020 dataset [18]. We found two observations that 1) ZSSR suffers from noise, and 2) the data augmentation methods, which are used in the original ZSSR, actually harm its SR performance (Table 3). Based on this observation, we decided to attach BM3D [4] before the ZSSR network optimization. For a fair comparison, we also use BM3D with our SimUSR. Regarding the data augmentation, we suspect that this is due to ZSSR network's small capacity and the severe spatial distortion by applying strong affine transformations [26].

By adding an ad-hoc denoiser (BM3D), ZSSR performance is dramatically improved by 0.63dB and 0.0422 in PSNR and SSIM, respectively. And by discarding affine augmentation, we can further enhance the ZSSR to achieve 26.55dB in PSNR (3rd row). With the same setting, our proposed SimUSR outperforms the ZSSR in a huge margin. For example, SimUSR with the lightweight SR network, CARN [2], already boosts the SR performance of the ZSSR by 0.64dB and 0.0176 in PSNR and SSIM, respectively. Moreover, thanks to the high flexibility of our method, we can easily improve the performance by simply changing the backbone to any other SR network. For instance, we get another 0.09dB improvement in PSNR by just replacing a backbone network from CARN to RCAN [29]. Figure 4 shows the qualitative comparison between the ZSSR and our method with different backbone networks. Similar to the bicubic SR task, SimUSR (with any backbone) provides better restoration results across various cases.

Table 4. The number of the parameters and runtime comparison of 480×320 LR images with scale factor ×4.

Method	ZSSR	SimUSR (Ours)		
		CARN	EDSR	RCAN
# Params.	0.23M	1.14M	15.6M	43.2M
Runtime	300.83s	0.12s	1.93s	1.07s

5.3. Execution time

In this section, we evaluate and compare the latency of ZSSR and our SimUSR (Table 4). Note that we benchmark the runtime speed on the environment of NVIDIA TITAN X GPU by generating a 1080p SR image on scale factor ×4. Although ZSSR has only 0.23M parameters, it requires a huge amount of runtime (300.83s) since it has to perform optimization and inference at runtime. In contrast, our proposed SimUSR only takes less than two seconds (1.93s) even if we use a heavy SR network (EDSR) as a backbone model. Comparing to the ZSSR, our method is at least 155 times faster than the ZSSR and if we use a lightweight SR network (CARN), 2,500 times faster (0.12s vs. 300.83s).

To embed the SR method to the real application, it is obvious that both SR performance and the latency are important aspects (*e.g.* SR system for the streaming service). However, the above analysis reflects that although ZSSR has nice properties, which does not need an HR image, applying it to the real application is challenging because of its high latency. On the other hand, our approach can meet the criteria that real applications demand (on both the performance and speed) by taking advantage of supervised SR. In addition, if necessary, we can further reduce the latency by replacing the backbone to more lightweight network thanks to the flexibility of our method.

5.4. NTIRE 2020 super-resolution challenge

This work is proposed to participate to the NTIRE 2020 super-resolution challenge [18]. This challenge aims to develop an algorithm for the real-world SR task similar to the prior challenge in AIM 2019 [17]. However, unlike the previous challenge, there exist no LR and HR image pairs in the dataset akin to the setup that we experiment in Section 5.2. We submitted our SimUSR to the first track (image processing artifact), where the clean image is degraded by the unknown image artifact and downsample kernel. In this challenge, models are evaluated using the PSNR, SSIM and LPIPS [28] metrics.

Authorized licensed use limited to: EPFL LAUSANNE. Downloaded on October 02,2020 at 15:11:59 UTC from IEEE Xplore. Restrictions apply.

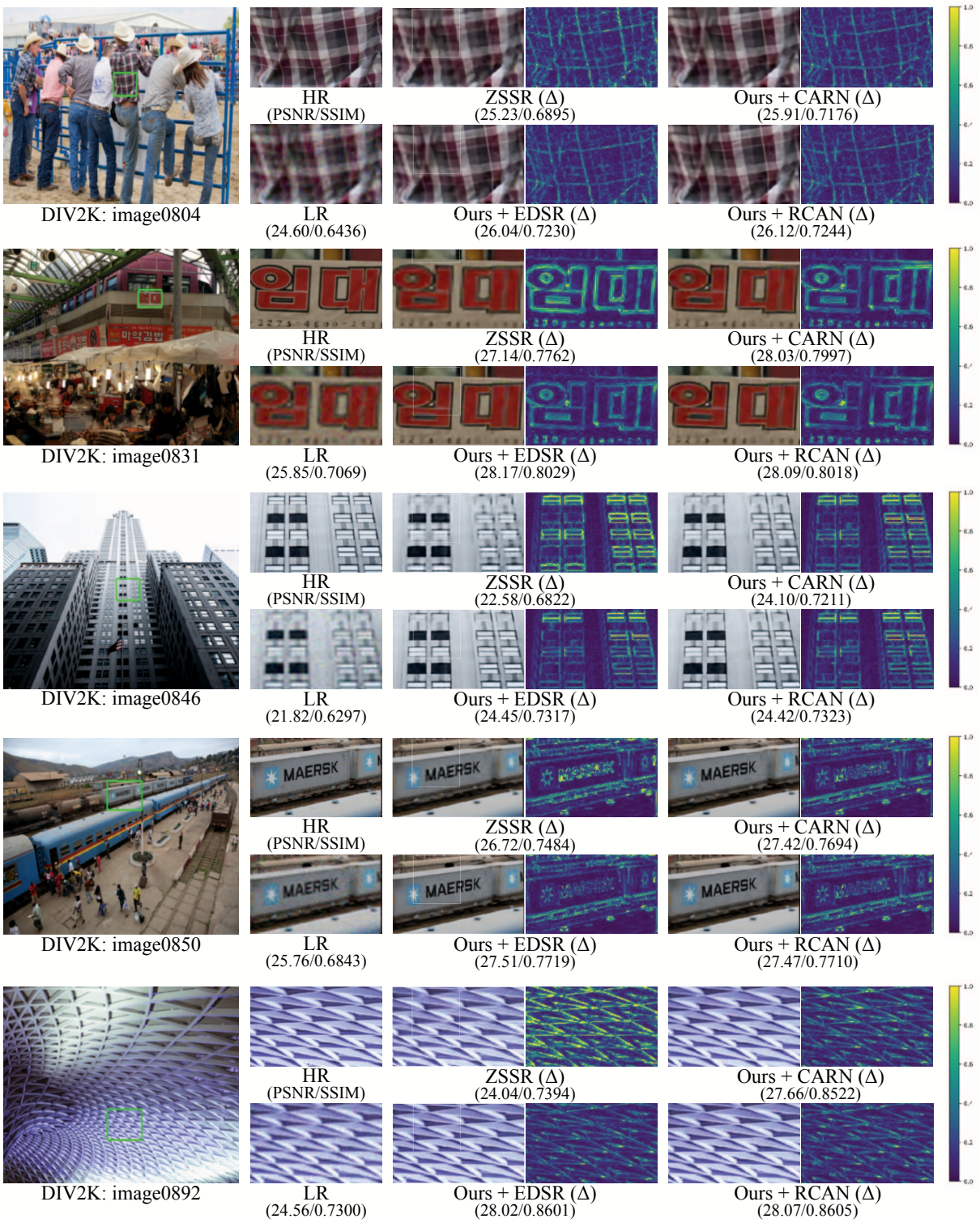

Figure 4. Qualitative comparison of using our proposed method on the NTIRE 2020 dataset [18]. Δ is the absolute residual intensity map between the network output and the ground-truth HR image.

Authorized licensed use limited to: EPFL LAUSANNE. Downloaded on October 02,2020 at 15:11:59 UTC from IEEE Xplore. Restrictions apply.

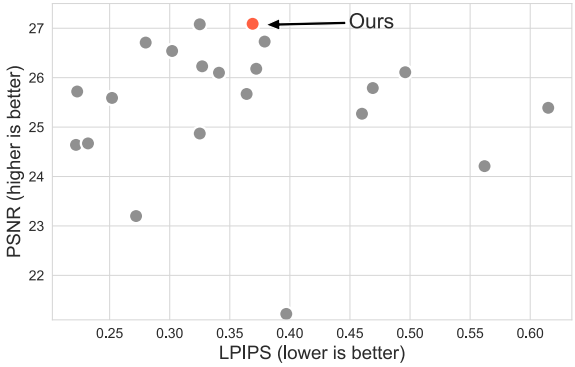

Figure 5. Performance comparison of each entry in the NTIRE 2020 super-resolution challenge [18] (track one). Our proposed SimUSR is marked as a red circle. Our method achieves the best PSNR with a reasonable LPIPS (13th rank).

Final result on the test dataset is shown in Figure 5 and Table 5. As shown in Figure 5, our method achieves the best PSNR score among all entries with a reasonable LPIPS score. Note that since we directly optimize the network using pixel-based loss, the LPIPS score of our SimUSR is lower than the rank of PSNR. We also report the challenge result sorted on the PSNR (Table 5). We get the best PSNR and second-best on SSIM with the 13th rank of LPIPS.

6. Discussion

In this section we discuss about the limitation of our method and the future direction.

Limitation. Though the accessibility to multiple LR images is a mild and reasonable relaxation in many cases, there are still many applications and domains that cannot resort on such assumption where collecting the data is very expensive, *e.g.*, medical imaging. In addition, SimUSR heavily relies on the generalizability of a model over different scales and pixel distributions, which can cause unexpected artifacts [26]. Because SimUSR uses bicubic downsampling to prepare the pseudo pairs, this may also cause an implicit bias in the SR model during the training. Last but not least, it is true that SimUSR is a basic approach that one would easily come up with but overlooked until now. We argue that it should be by no means a new state-of-the-art but serve as a reasonable baseline to beat in the future.

Future work. We showed that our SimUSR framework is a strong baseline but it still has a plenty of room to improve its performance. For example, we used the BM3D as the preprocessing module for removing the noise. This pre-module can be replaced to more effective models [13].

Table 5. Performance comparison of each entry in the NTIRE 2020 super-resolution challenge [18] (track one) sorted on the PSNR. The number in the parenthesis denotes the rank. Our proposed SimUSR is ranked the best in PSNR and second-best in SSIM.

Method	PSNR ↑	SSIM ↑	LPIPS ↓
Ours	27.09 (1)	0.77 (2)	0.369 (13)
Anonymous 1	27.08 (2)	0.78 (1)	0.325 (8)
Anonymous 2	26.73 (3)	0.75 (5)	0.379 (15)
Anonymous 3	26.71 (4)	0.76 (4)	0.280 (6)
Anonymous 4	26.54 (5)	0.75 (8)	0.302 (7)
Anonymous 5	26.23 (6)	0.75 (7)	0.327 (10)

7. Conclusion

We have introduced a simple but effective baseline for a fully unsupervised super-resolution task (SimUSR). we first clarify that unsupervised SR should strictly denote the task without any access to HR images. While complying with this definition, we assume that low resolution (LR) images are relatively easy to obtain in the real-world. Exploiting multiple LR images, we generated a pseudo-pair dataset of LR images and their down-scaled version and use this to train a SR model. This simple conversion allows us to enjoy the advantages of supervised learning. We demonstrated that our SimUSR outperforms previous unsupervised SR method while having very short latency. Moreover, by integrating the recently developed SR architectures and techniques, we showed that SimUSR successfully close the performance gap between the unsupervised and the supervised SR methods. Though our approach is simple, we argue that accessibility to multiple LR images is a legitimate setting and SimUSR serves as a strong baseline of unsupervised SR in this regime, which should be investigated prior to considering other complicated methods.

Acknowledgement. This work was supported by NAVER Corporation and also by the National Research Foundation of Korea grant funded by the Korea government (MSIT) (no.NRF-2019R1A2C1006608)

References

[1] Eirikur Agustsson and Radu Timofte. Ntire 2017 challenge on single image super-resolution: Dataset and study. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 126–135, 2017. 4

[2] Namhyuk Ahn, Byungkon Kang, and Kyung-Ah Sohn. Fast, accurate, and lightweight super-resolution with cascading residual network. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 252–268, 2018. 1, 2, 4, 6

[3] Marco Bevilacqua, Aline Roumy, Christine Guillemot, and Marie-Line Alberi-Morel. Low-complexity single-image super-resolution based on nonnegative neighbor embedding.

Authorized licensed use limited to: EPFL LAUSANNE. Downloaded on October 02,2020 at 15:11:59 UTC from IEEE Xplore. Restrictions apply.

In *Proceedings of the British Machine Vision Conference (BMVC)*, 2012. 4

[4] Kostadin Dabov, Alessandro Foi, Vladimir Katkovnik, and Karen Egiazarian. Image denoising by sparse 3-d transform-domain collaborative filtering. *IEEE Transactions on image processing*, 16(8):2080–2095, 2007. 1, 2, 4, 6

[5] Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang. Image super-resolution using deep convolutional networks. *IEEE transactions on pattern analysis and machine intelligence*, 38(2):295–307, 2015. 1, 2

[6] Ruicheng Feng, Jinjin Gu, Yu Qiao, and Chao Dong. Suppressing model overfitting for image super-resolution networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 0–0, 2019. 2

[7] Manuel Fritsche, Shuhang Gu, and Radu Timofte. Frequency separation for real-world super-resolution. *arXiv preprint arXiv:1911.07850*, 2019. 1, 2

[8] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In *Advances in neural information processing systems*, pages 2672–2680, 2014. 2

[9] Jinjin Gu, Hannan Lu, Wangmeng Zuo, and Chao Dong. Blind super-resolution with iterative kernel correction. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1604–1613, 2019. 2

[10] Reinhard Heckel and Paul Hand. Deep decoder: Concise image representations from untrained non-convolutional networks. *International Conference on Learning Representations*, 2019. 2

[11] Jia-Bin Huang, Abhishek Singh, and Narendra Ahuja. Single image super-resolution from transformed self-exemplars. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 5197–5206, 2015. 2, 4

[12] Jiwon Kim, Jung Kwon Lee, and Kyoung Mu Lee. Accurate image super-resolution using very deep convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1646–1654, 2016. 1, 2

[13] Alexander Krull, Tim-Oliver Buchholz, and Florian Jug. Noise2void-learning denoising from single noisy images. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2129–2137, 2019. 8

[14] Christian Ledig, Lucas Theis, Ferenc Huszár, Jose Caballero, Andrew Cunningham, Alejandro Acosta, Andrew Aitken, Alykhan Tejani, Johannes Totz, Zehan Wang, et al. Photo-realistic single image super-resolution using a generative adversarial network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4681–4690, 2017. 1

[15] Bee Lim, Sanghyun Son, Heewon Kim, Seungjun Nah, and Kyoung Mu Lee. Enhanced deep residual networks for single image super-resolution. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pages 136–144, 2017. 1, 2, 4

[16] Andreas Lugmayr, Martin Danelljan, and Radu Timofte. Unsupervised learning for real-world super-resolution. *arXiv preprint arXiv:1909.09629*, 2019. 1, 2

[17] Andreas Lugmayr, Martin Danelljan, Radu Timofte, et al. Aim 2019 challenge on real-world image super-resolution: Methods and results. In *ICCV Workshops*, 2019. 6

[18] Andreas Lugmayr, Martin Danelljan, Radu Timofte, et al. Ntire 2020 challenge on real-world image super-resolution: Methods and results. *CVPR Workshops*, 2020. 1, 4, 6, 7, 8

[19] David Martin, Charless Fowlkes, Doron Tal, and Jitendra Malik. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In *Proceedings of the International Conference on Computer Vision (ICCV)*, 2001. 4

[20] Yusuke Matsui, Kota Ito, Yuji Aramaki, Azuma Fujimoto, Toru Ogawa, Toshihiko Yamasaki, and Kiyoharu Aizawa. Sketch-based manga retrieval using manga109 dataset. *Multimedia Tools and Applications*, 76(20):21811–21838, 2017. 4

[21] Assaf Shocher, Nadav Cohen, and Michal Irani. "zero-shot" super-resolution using deep internal learning. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3118–3126, 2018. 1, 2, 3, 4, 6

[22] Jae Woong Soh, Sunwoo Cho, and Nam Ik Cho. Meta-transfer learning for zero-shot super-resolution. *arXiv preprint arXiv:2002.12213*, 2020. 3, 4

[23] Radu Timofte, Vincent De Smet, and Luc Van Gool. A+: Adjusted anchored neighborhood regression for fast super-resolution. In *Asian conference on computer vision*, pages 111–126. Springer, 2014. 1, 2

[24] Dmitry Ulyanov, Andrea Vedaldi, and Victor Lempitsky. Deep image prior. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 9446–9454, 2018. 2

[25] Jianchao Yang, John Wright, Thomas S Huang, and Yi Ma. Image super-resolution via sparse representation. *IEEE transactions on image processing*, 19(11):2861–2873, 2010. 4

[26] Jaejun Yoo, Namhyuk Ahn, and Kyung-Ah Sohn. Rethinking data augmentation for image super-resolution: A comprehensive analysis and a new strategy. *arXiv preprint arXiv:2004.00448*, 2020. 2, 4, 6, 8

[27] Yuan Yuan, Siyuan Liu, Jiawei Zhang, Yongbing Zhang, Chao Dong, and Liang Lin. Unsupervised image super-resolution using cycle-in-cycle generative adversarial networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 701–710, 2018. 1, 2

[28] Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 586–595, 2018. 6

[29] Yulun Zhang, Kunpeng Li, Kai Li, Lichen Wang, Bineng Zhong, and Yun Fu. Image super-resolution using very deep residual channel attention networks. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 286–301, 2018. 1 ,2 ,4 ,6

Authorized licensed use limited to: EPFL LAUSANNE. Downloaded on October 02,2020 at 15:11:59 UTC from IEEE Xplore. Restrictions apply.