School of Informatics

Informatics Research Review
Single Image Super Resolution: Finding the Correct Training Data

[redacted]

January 2021

Abstract

To train a super-resolution model, it is usually required to introduce training data in the form of high-resolution – low-resolution image pairs. There exists a problem that the image degradation model captured by the training data does not match those in real-life scenarios. This problem is often overlooked by some researchers and can result in overconfidence when applying the model on real-world data. In this review, we outline relevant researches and methods aiming to solve the above-mentioned problem.

Date: Friday 29th January, 2021

Supervisor: [redacted]

1 Introduction

Single image super-resolution (SISR) is the process of creating a high-resolution (HR) image from a single low-resolution (LR) image. Super-resolution has many potential applications where HR images might not be easily available due to restrictions in storage space, bandwidth, sensor quality etc. It can be beneficial in areas such as satellite imaging [1], medical imaging [2], surveillance [3], and camera quality enhancing [4].

Since deep learning technique is introduced to super-resolution in SRCNN [5], there has been increasing usage of it recently. While being able to generate convincing results with the pattern-capturing capability of neural networks, it also introduced a new problem. The goal of a super-resolution model is usually to learn a direct mapping from LR space to HR space, so it is common to use training data in the form of LR-HR image pairs. In earlier works, these pairs are generated by downsampling HR images to create perfectly aligned lower resolution images. Common downsampling methods include Gaussian kernel with sub-sampling [5, 6] and bicubic kernel [7, 8], sometimes with extra noises and filters applied. However, these downsampling methods are usually oversimplified, and the LR images produced by those might not be representative enough for real-world LR images. These models are tested on image pairs generated using the same methods, which means the evaluated performance will be falsely high and might struggle to generalize over real-world images. We refer to this as the image degradation problem.

In this review, we discuss some methods to mitigate the image degradation problem: improving the dataset, creating more representative LR images, and using real-world LR images directly. We also present a review of some well-performing researches in these methods and suggest some future work directions. We will ignore some other improvements and novelties introduced in these researches that are not directly related to our question.

2 Literature Review

In this section, we review some papers that have purposed methods to mitigate the inamge degradation problem. We compare the advantages and disadvantages of the purposed methods, and also point out some problems.

2.1 Specifically crafted datasets

When a dataset is faulty, the most straightforward solution is to abandon that dataset and find an appropriate alternative. In [9], Cai et al. aim to create a dataset (RealSR) with HR-LR image pairs which accurately capture the degradation model in real-world scenarios. In addition, the authors purposed a new network structure designed specifically for some unique properties in the RealSR dataset, but this will not be discussed in detail in this review.

The images in the RealSR dataset are captured by two full-frame DSLR cameras: a Canon 5D3 (resolution 5760×3840) and a Nikon D810 (resolution 7360×4912). The different resolution images of the same region are created by changing the focal length f of the cameras. A large f will result in a more zoomed-in picture which is able to capture more details. The scale of the images and the relative f are listed in table 1. With this method, images can be captured easily. A total of 595 image pairs of 234 scenes were captured in the RealSR dataset.

After capturing the images comes the difficult task of aligning them. First, the images will go

Scale	$\times 1$ (ground-truth HR)	$\times 2$	$\times 3$	$\times 4$
f	105mm	50mm	35mm	28mm

Table 1: Image scale and focal length used in [9]

Method	bicubic degradation	multiple degradations	RealSR
VDSR	32.63	32.65	33.64
SRResNet	32.66	32.69	33.69
RCAN	32.91	32.91	33.87

Table 2: Average PSNR (dB) of different methods and training sets, $\times 2$ scaling [9]

through a distortion correction algorithm and their central region will be cropped out and used in further procedures, denoted as I_H for the HR image and I_L for the LR image. At this step, I_L contains a larger area compared to I_H. Then, I_H will be used as a reference to crop out an aligned region of I_L, forming the final HR-LR pair. The cropping process is done by minimizing the following function from [9]:

$$\min_{\tau} ||\alpha C(\tau \circ I_L) + \beta - I_H||_p^p$$

where C is a cropping operation, τ is an affine transformation matrix, α and β are luminance adjustment parameters, and $|| \dots ||_p$ is an L_p norm. Then τ can be solved iteratively by a locally linear approximation. After solving for τ, we have the aligned LR image as $I_L^A = \alpha C(\tau \circ I_L) + \beta$ [9]. Taken with different focal length, the HR and LR images' scale and luminance condition might be different. With this purposed method, the authors were able to get pixel-wise accuracy, while other popular algorithms such as SURF [10] and SIFT [11] fails.

In this research, the authors tested the improvements brought by the RealSR dataset. They trained three neural networks: SRResNet [7], VDSR [12], and RCAN [13], each on three datasets: the DIV2K [14] dataset with bicubic degradation, DIV2K with multiple degradations [15], and RealSR. After training, they are applied to the RealSR testing set. The performance is evaluated using the average peak signal-to-noise ratio (PSNR) and structural similarity index measure (SSIM). With the super-resolution scale of $\times 2$, $\times 3$, and $\times 4$, and for all three networks, the models trained on the RealSR dataset outperforms those trained on traditionally generated LR images. The results also showed that using multiple degradations can improve generalization comparing to bicubic degradation, but only by a small margin.

Although the results seem promising, there lie some critical problems in this research. In the RealSR dataset, images of different resolution are captured by zooming a DSLR camera. This can be better than images acquired by bicubic degradation, or multiple degradations, but it ultimately faces the same problem: this is only a small portion of real-world LR images. In application, LR images could be affected by many factors such as depth of field, digital compression, even motion blur. So models trained using RealSR still cannot be treated as if it generalizes universally.

Also, the images are captured with different focal length, which means the blur kernel varies each scene and the depth of field changes with the focal length. This problem is pointed out in this research and a special network (the Laplacian pyramid based kernel prediction network) is created to address it. This means that the RealSR dataset is not fully compatible with other popular network architectures. Define a magical function $I_H^P = \mathcal{M}(I_L)$ which creates a perfect

HR image I_H^P, RealSR will have $I_H = \mathcal{M}(I_L) + noise$, which means any models trained on it will always be a bit off.

Another issue in this paper is that the evaluation strategy might not be fair. The networks are not trained on the same data: the two networks meant to demonstrate the inferior traditional degradation methods are trained on DIV2K, while the third network is trained on RealSR, but all of them are tested on the RealSR testing set. Arguably, the RealSR training set will have many similarities with the testing set, while images from DIV2K will be far less similar. This gave the network trained on RealSR an unfair advantage which makes the PSNR obtained inaccurate, and it means the improvement from using RealSR might not be as big as it shows. A quick improvement could be using the same HR images from the RealSR dataset, but downsampling them with traditional degradation methods to generate the LR images for the two networks currently trained on DIV2K.

With all being said, these specifically captured datasets could still be beneficial and bring an improvement to existing methods. One example of this would be [4], where the super-resolution model only aims to enhance one specific type of image. However, although the image-capturing process could be further automated, creating such a dataset still includes lots of work and could be expensive to scale, so alternatives are desired.

2.2 Image degradation with deep learning

Since good datasets are rare and can be hard to create, another method is to learn a good degradation model that captures how real-world LR images are formed. Training such a model using traditional deep learning would require the same HR-LR pair which cannot be acquired. This makes learning the degradation pattern a difficult task.

In [16], Bulat et al. purposed a method which uses GAN to learn the degradation model. This method does not require any paired data, instead, it only requires two disjoint datasets: one for HR images and one for LR images, both are easy to acquire. The complete architecture contains two GANs: an HR-LR network and an LR-HR network.

First, images from the HR image dataset will be fed into the HR-LR network, which consists of an HR-LR generator and an LR discriminator. The goal of this network is to convert the inputted HR image to a realistic LR image while learning the degradation model during this process. In real-world, an HR image could have many different LR images, due to the many types of noise that could have an effect on it. To simulate this behaviour, the HR image will be concatenated with a random noise vector, then used as the input of the generator, and be converted to an LR image. Then the discriminator which have access to the LR image dataset will attempt to classify the generated LR image as real (authentic) or fake (generated). Using this method, we are able to generate LR images that are more similar to real-world LR images.

The LR-HR network is a common GAN network used in super-resolution. It has a generator which converts LR images to HR images. The LR image inputs for this network is generated by the HR-LR network, and the generated HR image will be given to a discriminator which classifies HR images as real or fake.

The loss function for each network is defined as a combination of L_2 pixel loss and GAN loss [16]:

$$\mathcal{L} = \alpha \mathcal{L}_{pixel} + \beta \mathcal{L}_{GAN}$$

where α and β are weights. It is noteworthy that in previous work, the loss is mostly driven by the pixel loss, which focuses on how the generated image is related to the original image. In

Method	FID	PSNR
	LR test set	LS3D-W
SRGAN	104.80	23.19
LR-HR network	85.59	23.50
CycleGan	19.01	16.10
Purposed method	14.89	19.30

Table 3: FID-based evaluation on real-world LR test set and PSNR-based evaluation on LS3D-W [16]

[16], the loss is GAN-centered ($\beta\mathcal{L}_{GAN} > \alpha\mathcal{L}_{pixel}$), which focuses on how the generated image is related to other real-world images of the same type. This also helps in generating realistic LR images.

In this research, the authors compared the performance of the purposed method and several other well-performing methods for image super-resolution as well as face super-resolution. Here we will only list some of them, namely, SRGAN [7], CycleGan [17], the aforementioned LR-HR network, and the complete purposed method. The CycleGan is trained similarly with the purposed method using unpaired images, and the other two methods are trained with paired bilinearly downsampled HR images.

The trained networks are evaluated on two testing sets. First, they are evaluated on a testing set consists of real-world LR images. Since there are no ground-truth HR images for this testing set, the performance is measured using Fréchet inception distance (FID). They are also evaluated on 1000 images from the LS3D-W dataset [18], with bilinearly downsampled images as the input, and the performance is measured using PSNR.

With the results, we can see that when tested with the bilinearly downsampled test data, SRGAN and LR-HR network had a better performance, but that is expected because they are specifically trained to work with images generated using this method. However, methods trained on unpaired data performs significantly better on the real-world LR images. This proves that using simple methods to generate LR images does affect the network's generalizing ability, and the method purposed in this research does provide a huge improvement. Unlike the evaluation described in [9], this evaluation does not give an unfair advantage to any method, because all methods have access to images that are similar to the testing set during training.

Although it showed a huge improvement on real-world LR images, this method is not without

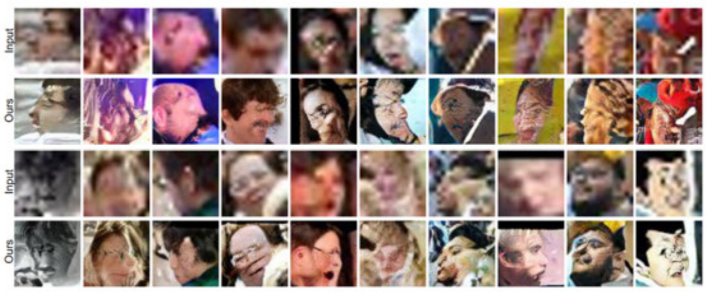

Figure 1: Examples of failure cases generated by GAN. [16]

its flaws. As mentioned above, this method is GAN-driven, which means the lack of direct supervision makes it focuses on fooling the discriminator rather than looking like the input. In this research, the authors reported a 10% fail rate where the generated HR image is not realistic at all. This is most likely because the generated image is compared with an entire HR dataset rather than just the input image. It is a fairly high fail rate, combined with the horrifying image it produced, the method is not very practical. Also, Bulat et al. focused on the super-resolution of facial images. This is different from other images because we can easily learn some critical prior knowledge about facial features, but this is not the case with images of objects and sceneries.

While Bulat et al. focused directly on learning the image degradation process, Lugmayr et al. had a different approach. In [19], they argued that the commonly used bicubic degradation will drastically change and remove many features in an image such as sensor noise and compression artifacts. This results in very "clean" images and networks trained on it will not be able to generalize to more noisy real-world LR images. In this research, the authors purposed a method to restore those lost features to the generated LR image.

The network purposed could be separated into two modules: a domain-correction network and a super-resolution network. The super-resolution network could be using any purposed architecture using paired data, and will not be covered in this review. The two modules are trained separately to prevent the loss of the super-resolution network from affecting the domain-correction network.

The task of the domain-correction network is to map bicubic degraded images to the "real-world" images distribution. Firstly, a generator will be employed to convert the input bicubic degraded image to one that can pass the discriminator which has access to real-world LR images. To make sure the first GAN attempts to preserve as much image content as possible, another set of generator and discriminator will be employed to map the generated image back to the bicubic degraded images distribution. The two generators are "connected" with the cycle consistency loss [17]. The loss of the domain-correction network is defined as [19]:

$$\mathcal{L} = \mathcal{L}_{GAN1} + \mathcal{L}_{GAN2} + \lambda\mathcal{L}_{cyc}$$

where λ is a weight used to balance the cycle consistency loss. Then the super-resolution network can be trained with full supervision. For evaluating the results, this method did not have a higher statistical score when compared to other methods but have achieved the best perceptual score in both sensor noise and JPEG artifacts. This network eliminates the problems that [16] had. It is trained on sceneries rather than faces, and the loss when training the super-resolution network is mostly driven by pixel-wise losses rather than the GAN loss. This is one of the best methods when it comes to real-world super-resolution.

2.3 Unsupervised image super-resolution

Although the aforementioned methods do not require paired data, the super-resolution networks are still trained in a supervised manner. In CinCGAN [20], Yuan et al. purposed a completely unsupervised method for image super-resolution. The structure of CinCGAN can be viewed as two nested CycleGANs. The first CycleGAN maps real LR images to clean LR, and maps it back to the original real LR to preserve content. The purpose of this CycleGAN is to deblur and denoise the LR images. Then the processed clean LR images are mapped to HR images in the second cycleGAN to increase the resolution. In this method, no LR-HR pair is ever used. Comparing this method to the one purposed by Lugmayr et al., we see one critical difference:

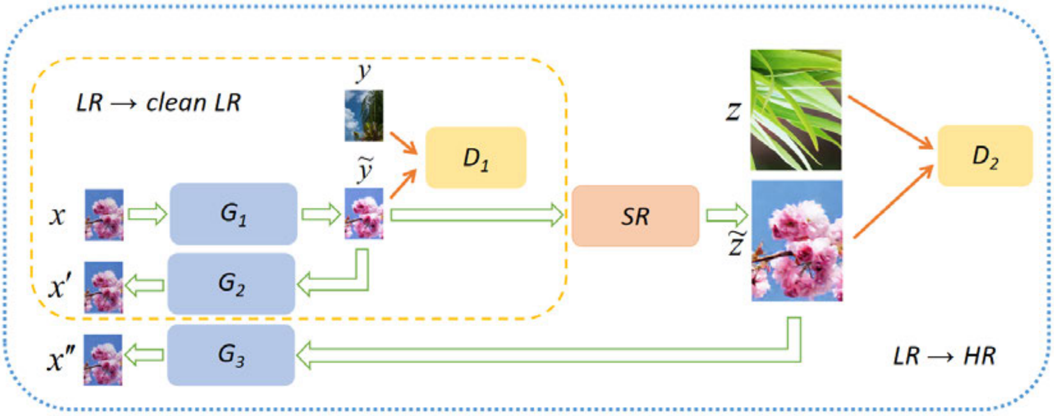

Figure 2: The nested cycleGAN structure of CinCGAN. [20]

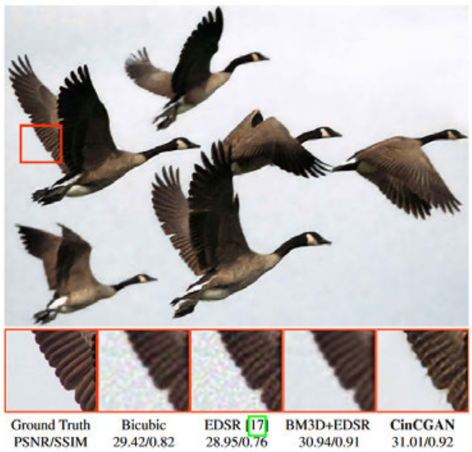

Figure 3: Super-resolution results with a noisy input. [20]

Lugmayr et al. attempts to map the degraded images to real-world images in order to train a super-resolution network for real-world images, where Yuan et al. attempts to map the real-world images to a clean image similar to the bicubic degraded images. This means during inference, the approach of Yuan et al. will have extra steps cleaning the input image.

In section 2.2, the methods reviewed trains the super-resolution network with smartly generated images. This purposed methods instead trains the super-resolution network with real-world LR images directly. This means that this method is more "familiar" with real LR images and is potentially more robust. Although the super-resolution network does not have supervision from an HR image, the image content is still preserved by the usage of the CycleGAN structure. However, the lack of pixel-wise loss does means training will be harder.

Also, this work focuses on cleaning the input image so it is more similar to a bicubic downsampled image. However, according to Lugmayr et al. [19], the lack of robustness is caused by the loss of information when applying bicubic degradation. If this is the case, then mapping real images to bicubic degraded images might not have a result as good as the other option, because

Figure 4: Performance comparison between SRGAN and ZSSR. [21]

the super-resolution network is still working on images similar to bicubic degraded images.

Next, we would like to introduce a different approach to unsupervised learning: deep internal learning. In [21], Shocher et al. purposed ZSSR which learns to super-resolute an image from the image itself. The logic behind this is that small pieces of information within an image usually repeats and scales across the image many times [22], so an image contains enough information to enhance itself within itself, and based on the idea that the visual entropy inside a single image is much smaller than the visual entropy of a collection of images [22], Shocher et al. believed that it should be feasible to train an image-specific network to perform super-resolution on that image. For training the image-specific network, the input image is downsampled using a bicubic downscaling kernel, then the downscaled image will be used together with the original image to crop out training pairs which are used to train the image-specific network. Learned from the input image itself, it is much more specific and can achieve better performance even when the downsampling method is unknown.

As an addition to ZSSR, KernelGAN [23] is proposed. While ZSSR uses a bicubic kernel to scale the images, KernelGAN uses a GAN to learn the downscaling kernel of the image. The generator will downscale the image using a deep linear network, then a small image will be cropped out from the downscaled image and the original image. The cropped images will be used to train the discriminator, which outputs a heatmap representing for each pixel, how likely its surroundings are from the original image. After training, we can get a learned downscaling kernel from the generator, and this kernel can be used by ZSSR and other similar designs.

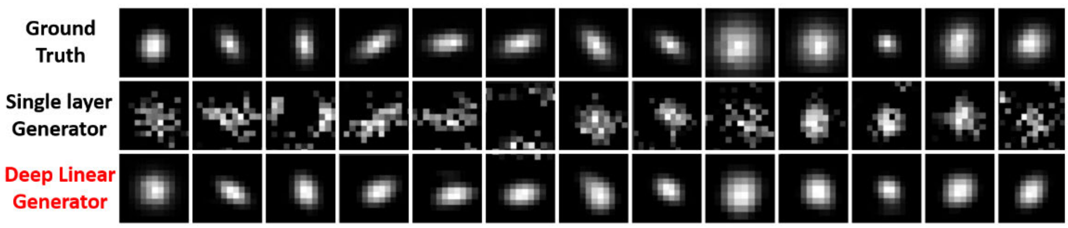

Figure 5: Downscaling kernel learned by deep linear generators. [23]

The advantage of unsupervised learning is obvious: it does not require paired data, so we do not have to generate paired data to train it, unlike methods reviewed in section 2.2. However, unsupervised training methods are usually vastly different from supervised learning methods, so unsupervised super-resolution has its own research circle. On the contrary, if we can learn to generate good paired data, then the technique can be applied to all other supervised super-resolution methods and improve their performance.

2.4 A side note: Style transfer in image degradation

As a noteworthy discovery, many methods above are heavily inspired by CycleGAN [17]. Interestingly, it is not about super-resolution, but about image-to-image translation. CycleGAN is an improvement on pix2pix [24] which requires paired data. It is able to train on two unpaired datasets, and map images from one dataset to the distribution of another dataset. This is usually known as style transfer and is commonly used to generate images which have the same context but have a different style, e.g. photo to painting, and semantic segmentation to photo. We can see that this technique is used in the above-reviewed researches during LR image generation and even HR image generation. This shows that "generated" and "real", "low-resolution" and "high-resolution" can also be considered to be two styles to some extent, and style transfer might be applicable in much more areas than one might expect.

3 Summary & Conclusion

In this review, we reviewed some methods to improve the generalizing ability of super-resolution methods by switching out the simple bicubic degraded images. Capturing those low-resolution high-resolution images pairs by camera is an easy option to think of but is the least preferable one in our view. It requires lots of extra work, can be very expensive on a larger scale, and are often also limited in variations of degradation factors. The better option would be using GANs to assist in degrading images in an unsupervised manner. The training data are easy to acquire, it can generate good paired data which can be used on super-resolution network and provide pixel-wise supervision, and it can easily be scaled to fit a large dataset. Unsupervised methods do have the advantage of being able to be trained directly on real LR images, but this advantage can be minimized by learning better degradation models, and they lack the ability to adapt to well-developed supervised training methods.

It is shown that using simple bicubic degradation to generate training data does affect the robustness of the network, and using the aforementioned method can improve the perceptual score of super-resolution. Having an effective solution to our initial problem, we believe the future direction could be task-specific degradation, where each degradation model can focus on images from a narrower domain e.g. being the same type of domain or sharing the same type of noise, or even having a different one for every single image. Then these degradation models can work simultaneously and/or sequentially to create a dataset, which could be used for future research and evaluation.

Also, Ledig et al. [7] mentioned a problem with the commonly-used PSNR evaluation metric. Not only it cannot be used to evaluate real-world LR images, sometimes achieving high PSNR could make the generated image lack high-frequency content, resulting in a lower perceptual score. One future direction could also be developing an evaluation metric that can estimate the perceptual score, so it can be used on real-world LR images with no ground-truth.

References

[1] Jacob Shermeyer and Adam Van Etten. The effects of super-resolution on object detection performance in satellite imagery. In *IEEE Conference on Computer Vision and Pattern Recognition Workshops, CVPR Workshops 2019, Long Beach, CA, USA, June 16-20, 2019*, pages 1432–1441. Computer Vision Foundation / IEEE, 2019.

[2] Yuhua Chen, Feng Shi, Anthony G. Christodoulou, Yibin Xie, Zhengwei Zhou, and Debiao Li. Efficient and accurate MRI super-resolution using a generative adversarial network and 3d multi-level densely connected network. In Alejandro F. Frangi, Julia A. Schnabel, Christos Davatzikos, Carlos Alberola-López, and Gabor Fichtinger, editors, *Medical Image Computing and Computer Assisted Intervention - MICCAI 2018 - 21st International Conference, Granada, Spain, September 16-20, 2018, Proceedings, Part I*, volume 11070 of *Lecture Notes in Computer Science*, pages 91–99. Springer, 2018.

[3] Liangpei Zhang, Hongyan Zhang, Huanfeng Shen, and Pingxiang Li. A super-resolution reconstruction algorithm for surveillance images. *Signal Process.*, 90(3):848–859, 2010.

[4] Xuaner Zhang, Qifeng Chen, Ren Ng, and Vladlen Koltun. Zoom to learn, learn to zoom. In *IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2019, Long Beach, CA, USA, June 16-20, 2019*, pages 3762–3770. Computer Vision Foundation / IEEE, 2019.

[5] Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang. Learning a deep convolutional network for image super-resolution. In *European conference on computer vision*, pages 184–199. Springer, 2014.

[6] Wenzhe Shi, Jose Caballero, Ferenc Huszár, Johannes Totz, Andrew Aitken, Rob Bishop, Daniel Rueckert, and Zehan Wang. Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. 06 2016.

[7] Christian Ledig, Lucas Theis, Ferenc Huszar, Jose Caballero, Andrew Cunningham, Alejandro Acosta, Andrew P. Aitken, Alykhan Tejani, Johannes Totz, Zehan Wang, and Wenzhe Shi. Photo-realistic single image super-resolution using a generative adversarial network. In *2017 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, Honolulu, HI, USA, July 21-26, 2017*, pages 105–114. IEEE Computer Society, 2017.

[8] T. Tong, G. Li, X. Liu, and Q. Gao. Image super-resolution using dense skip connections. In *2017 IEEE International Conference on Computer Vision (ICCV)*, pages 4809–4817, 2017.

[9] Jianrui Cai, Hui Zeng, Hongwei Yong, Zisheng Cao, and Lei Zhang. Toward real-world single image super-resolution: A new benchmark and a new model. *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, 2019.

[10] Herbert Bay, Tinne Tuytelaars, and Luc Van Gool. Surf: Speeded up robust features. In *European conference on computer vision*, pages 404–417. Springer, 2006.

[11] David G Lowe. Distinctive image features from scale-invariant keypoints. *International journal of computer vision*, 60(2):91–110, 2004.

[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.

[13] 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.

[14] Eirikur Agustsson and Radu Timofte. Ntire 2017 challenge on single image super-resolution: Dataset and study. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, July 2017.

[15] K. Zhang, W. Zuo, and L. Zhang. Learning a single convolutional super-resolution network for multiple degradations. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3262–3271, 2018.

[16] Adrian Bulat, Jing Yang, and Georgios Tzimiropoulos. To learn image super-resolution, use a gan to learn how to do image degradation first. In *Proceedings of the European conference on computer vision (ECCV)*, pages 185–200, 2018.

[17] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proceedings of the IEEE international conference on computer vision*, pages 2223–2232, 2017.

[18] Adrian Bulat and Georgios Tzimiropoulos. How far are we from solving the 2d & 3d face alignment problem?(and a dataset of 230,000 3d facial landmarks). In *Proceedings of the IEEE International Conference on Computer Vision*, pages 1021–1030, 2017.

[19] Andreas Lugmayr, Martin Danelljan, and Radu Timofte. Unsupervised learning for real-world super-resolution. In *2019 IEEE/CVF International Conference on Computer Vision Workshops, ICCV Workshops 2019, Seoul, Korea (South), October 27-28, 2019*, pages 3408–3416. IEEE, 2019.

[20] 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 *2018 IEEE Conference on Computer Vision and Pattern Recognition Workshops, CVPR Workshops 2018, Salt Lake City, UT, USA, June 18-22, 2018*, pages 701–710. IEEE Computer Society, 2018.

[21] Assaf Shocher, Nadav Cohen, and Michal Irani. "zero-shot" super-resolution using deep internal learning. In *2018 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2018, Salt Lake City, UT, USA, June 18-22, 2018*, pages 3118–3126. IEEE Computer Society, 2018.

[22] Maria Zontak and Michal Irani. Internal statistics of a single natural image. In *The 24th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2011, Colorado Springs, CO, USA, 20-25 June 2011*, pages 977–984. IEEE Computer Society, 2011.

[23] Sefi Bell-Kligler, Assaf Shocher, and Michal Irani. Blind super-resolution kernel estimation using an internal-gan. In Hanna M. Wallach, Hugo Larochelle, Alina Beygelzimer, Florence d'Alché-Buc, Emily B. Fox, and Roman Garnett, editors, *Advances in Neural Information Processing Systems 32: Annual Conference on Neural Information Processing Systems 2019, NeurIPS 2019, December 8-14, 2019, Vancouver, BC, Canada*, pages 284–293, 2019.

[24] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A. Efros. Image-to-image translation with conditional adversarial networks. In *2017 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, Honolulu, HI, USA, July 21-26, 2017*, pages 5967–5976. IEEE Computer Society, 2017.