International Journal of Interdisciplinary Innovative Research &Development (IJIIRD)
ISSN: 2456-236X
Vol. 08 Special Issue 01 | 2023

A Comprehensive Survey of Single Image Super Resolution Techniques

Steve Dsouza[1], Yash Sanaye[2], Bajarang Londhe[3], Meena Ugale[4]

[1,2,3] *Student, Dept. of Information Technology, Xavier Institute of Engineering, Maharashtra, India*
[4] *Professor, Dept. of Information Technology, Xavier Institute of Engineering, Maharashtra, India*

ABSTRACT

Single image super-resolution (SISR) is a technique with the intent of restoring or recreating a high-resolution (HR) image from its corresponding low-resolution (LR) observation. The recent decade has seen immense research and development in image processing with particular attention to deep learning-based and Denoising Diffusion based image super-resolution (SR) approaches. Previous studies show excellent performance of existing image super-resolution techniques on synthetic data, but when faced with real-world images there is a drastic difference between the resulting performance, indicating an overestimation of their capability to super-resolve real-world images. To tackle this problem, numerous types of research have been conducted and researchers have come up with new approach strategies and techniques to solve the problem. The intent of this article is to make an extensive review of the state-of-the-art (SOTA) real world single image super-resolution (RSISR) methods and techniques along with the datasets used to train various models (skewing more towards face and textual SR). A detailed review of all the assessment metrics used and or considered to evaluate the results of the models along with their importance and specificity are discussed. The review mainly covers RISR/RSISR methods which include, Deep-learning techniques, image-pair based techniques, degradation-based methods, and Denoising Diffusion based RSISR techniques. Lastly, evaluation criteria and their significance along with further research directions are briefly discussed.

Keyword : - Deep Learning, Review, Real World Image Super Resolution, Super Resolution, Synthetic Super Resolution.

1. INTRODUCTION

With the rapid advancement in machine learning and AI, High-Resolution (HR) images have become an essential component for various image-based machine learning and AI paradigms. High Resolution (HR) images find applicability and a high level of desirability in many applications and domains pertaining to Intelligent surveillance, text recognition, segmentation, medical imaging, classifications, face enhancement, remote sensing, satellite imaging, aerospace, etc. There are in general two ways to acquire images with higher resolution. The first which natural intuition suggests is the use of modern image-capturing hardware or hardware upgradation. Although imaging devices and techniques have witnessed an obvious and significant progress in recent years, this approach has some profound limitations; 1) It is inflexible and economically taxing because the demand and requirements in practical application change constantly. 2) Enhancement of old low-resolution (LR) images is not possible, rather capturing new HR image data is all that can be accomplished. Rather than a hardware upgrade-based approach, the other solution, which is signal processing-based image resolution enhancement (i.e., software-based enhancement) is much more flexible and cost-efficient. Images with resolutions beyond the limits of existing low analog-based imaging systems can be obtained using SR techniques.

Image Super Resolution is a method developed to rebuild a realistic high resolution (HR) image from a corresponding low resolution (LR) image. Some of the proposed SISR techniques, which mainly include, "reconstruction-based" [19], example-based [20], sparse representation-based [14], denoising diffusion based and deep-learning models [2]- [7] etc., show amazing result on synthetic data and some of the newer models have even achieved commendable performance on real-world data (i.e., LR-HR conversion). Nevertheless, previous research has pointed out that numerous current SISR methods may overstate their super-resolution capability in reference to synthetic data that exhibit a considerable gap between realistic and synthetic data domains. To simplify, the SISR techniques may perform phenomenally well on synthetic data but when posed with real-world data they fail to show the same level of performance. To address this issue, researchers have modified their approach to target the problem of image super resolution in real world. This has led to a number of studies being conducted on building and collecting real-world datasets [14], assessment metrics and super resolution models for real world images.

This paper aims at providing a survey of recent and pertinent RSISR models and techniques, standard datasets, assessment metrics, and relevant studies, specifically Deep-learning based and Denoising diffusion Probabilistic models. There have been several works adhering to provide a study of SR technologies related to videos and images. For example, [14] has made a summary of the SISR methods, techniques, datasets and has provided a comprehensive review on some of the recent and existing deep-learning techniques and approaches trying to solve RSISR. This work however, is an attempt to extend the work with new relevant studies and approaches that have gained traction in recent years for RSISR such as denoising diffusion probabilistic models (diffusion models).

This study is structured as follows. Section 2 gives a brief background on RSISR. In Section 3, the datasets are described. In section 4 the various Assessment Metrics for assessing the super-resolved images on various factors are discussed. Section 5 reviews RSISR methods and techniques. Section 6 consists of Evaluation Criteria for the Super Resolution techniques. Lastly, Section 7 concludes the study.

2. BACKGROUND

SISR is the process of recovering a HR image from an LR observation. In general, it can be stated that after undergoing some degradation process the HR image gets converted to the LR image. It can be mathematically represented as,

$$Y = \phi(X, \theta) \tag{1}$$

Where Y is the observed LR image defined by ϕ which denotes the degradation process that X (HR original image) undergoes with certain parameters θ. θ includes model parameters such as blurring kernel, down-sampling operations, noise injection, compression artifacts, etc. It is important to acknowledge that the degradation process is unknown in the real world. The objective of SISR is to recover the most accurate estimate of the HR image from the observed LR image by reversing the degradation process, which can be represented as,

$$\hat{X} = \phi^{-1}(Y, \delta) = Z(Y, \delta) \tag{2}$$

Where Z is the associated with Z, $\hat{X}$ represents the super-resolved result. The processes pertaining reverse degradation process (inverse degradation model) and δ indicates the parameters to degradation and SR are apparently inverses of one another. Therefore, the SR function should be adapted to the degradation, for getting good reconstruction performance. In real world the degradation model and parameters are unknown, the only known element is the LR observation. So to tackle this, researches use mathematical models to simulate or generate LR-HR pairs to train their model. These mathematical models may include blur kernels, Gaussian noise injection, down sampling kernels etc.

Most SISR techniques and algorithms experience a significant performance drop and produce sub-par results when subjected to LR images from the real world. This is owing to the fact that the degradation process adopted by the SISR techniques fail to match the actual losses suffered by images in the real world. For this reason, some researchers have shifted their attention to look at the problem from different perspectives such as building realistic (real world) datasets, better performance metric analysis, SR model development, optimization advancements, alternative degradation estimation techniques etc.

3. DATASETS

Table -1: Datasets

Dataset	Published Year	Synthetic/ Realistic	Keywords
DRealSR	2022[22]	Realistic	Variational Focal Length
TextZoom	2020[7]	Realistic	Text, Recognition
CelebA-HQ	2018[17]	Realistic	High Resolution, Contain Noticeable Artifacts (aliasing, compression, blur)
Flickr-Faces-HQ	2019[3]	Realistic/ Semi-Synthetic	Variational age, ethnicity and background image, auto aligned and cropped

The datasets mentioned below are briefly summarized in Table -1:.

1) DRealSR [22] : Wei et al's [22] real-world dataset DRealSR is similar to RealSR but on a larger scale. In particular, five DLSR cameras from bands like Sony, Olympus, Canon, Panasonic and Nikon are used to take images from indoor and outdoor scenes at four different resolutions. To align the images with different resolutions, the "SIFT algorithm" is used. DRealSR [22] includes, "884, 783, and 840 LR-HR image pairs for the x2, x3, and x4 SR, respectively".
2) TextZoom [7]: Wang et al. developed a dataset called TextZoom specifically for the super-resolution of text images in real-world scenes. The dataset was created by combining images from two existing datasets, RealSR and SR-RAW [7], which comprise of natural scenes, including but not limited to shops, streets, and vehicle interiors. Text images in TextZoom were selected and annotated with details about their content, direction, and focal length. The dataset is divided into three levels of difficulty: easy, medium, and hard, and can be used for research on both text image super-resolution and text recognition due to its comprehensive annotations.
3) CelebA-HQ [17]: A dataset created by Karras et al. It comprises of "30,000 images of celebrity faces", sourced from the CelebA dataset and converted into high quality images at 1024x1024 resolution.
4) Flickr-Faces-HQ Dataset (FFHQ) [3]: T. Karras et al. have created FlickrFaces-HQ (FFHQ); a dataset of human faces that includes, "70,000 high-quality images at 1024x1024 resolution". The dataset exhibits greater diversity in terms of age, ethnicity, and image background compared to CELEBA-HQ. Moreover, it provides more extensive coverage of accessories, such as eyeglasses, sunglasses, hats, etc. The images were obtained from Flickr, inheriting its inherent biases, and were automatically aligned and cropped. Only images with permissive licenses were collected, and several automated filters were used to eliminate irrelevant content. Additionally, Mechanical Turk was utilized to remove any statues, paintings, or photos of photos that were deemed inappropriate for the study.

4. ASSESSMENT METRICS

Predominantly, two types of quality assessment methods are used to evaluate super-resolved images: subjective evaluation based on human perception and objective evaluation based on metrics. The former aligns better with real-world needs but has limitations such as personal preferences or bias affecting the assessment and being time-consuming and not easily automatable. It may be more convenient to use objective evaluation but the results may not be consistent with subjective evaluation [1]. Table -2 gives an overview of all the most commonly used standardized assessment metrics .

1) PSNR : Peak signal-to-noise ratio (PSNR) is a frequently and generally used metric for assessing the effectiveness and quality of image restoration techniques such as super-resolution (SR), denoising, deblocking, and deblurring. Given $\hat{X}$ and X, the PSNR is defined as,

$$PSNR = 10.log_{10}(\frac{L^2}{MSE}) \quad (3)$$

where $MSE = \frac{1}{HWC}\left\|X - \hat{X}^2\right\|_2^2$ denotes the mean squared error (MSE) between $\hat{X}$ and X, and L represents the maximum pixel value (i.e., 255 for 8-bit images). It can be seen from Eq. ((3)) that PSNR measures the proximity between corresponding pixels in the restored image $\hat{X}$ and the reference image X. In certain instances, the emphasis on discrepancies at the pixel level may not consistently align with perceptual quality.
2) SSIM [22]: Structure similarity index (SSIM) is a metric used to measure the structural similarity between two images. Compared to PSNR, this method is considered to more accurately capture the perceived visual quality of an image as it accounts for contrast, luminance and structure. SSIM can be denoted as,

$$SSIM = [S(I_{HR}, I_{SR})]^{\alpha} * [C(I_{HR}, I_{SR})]^{\beta} * [l(I_{HR}, I_{SR})\ ]^{\gamma} \quad (4)$$

Where, $S(I_{HR}, I_{SR})$, $C(I_{HR}, I_{SR})$ and $l(I_{HR}, I_{SR})$ express similarity in structure, contrast and luminance of the two images i.e. HR and SR image. While, α, β and γ represent the weighting parameters.
SSIM loss was specifically developed to enhance structural similarity between the HR image (ground truth) and the SR image. The SSIM loss function can be denoted as follows:

$$L_{SSIM}(I_{HR}, I_{SR}) = \frac{1}{2}(1 - F_{SSIM}(I_{HR}, I_{SR})) \tag{5}$$

where F_{SSIM} denotes the function of SSIM. When a reference image (also known as a ground truth image) is available, to evaluate the quality of the restored image PSNR and SSIM both can be used together.

Table -2: Assessment Metrics

Assessment Metric	Keywords
PSNR	Peak signal to noise ratio, Mean squared error
SSIM	Structural similarity, Luminance,
IFC	Natural scene statistics, Gaussian
LPIPS	Deep features, Human perceptual
PIQE	Perceptually significant spatial regions, Block level distortion map
NRQM	Statistical features, Regression forests, Linear regression model
MOS	Subjective assessment metric, Human perceptual score assignment, Arithmetic mean rating
FID	Diversity and Fidelity presence, Generative model, distributional comparison (generated image and ground truth)

3) IFC [23]: The information fidelity criterion (IFC) is a measure (metric) that assesses the excellence of images by considering natural scene statistics. Studies have demonstrated that the attributes of the space created by natural images can be described utilizing diverse models, such as the Gaussian scale mixtures. Generally, distortions in an image can disrupt these statistics and make the image appear unnatural. The mutual information between a test image and a reference image is evaluated by IFC, which utilizes natural scene and distortion models. Research and various studies have indicated that IFC is effective when evaluating the quality of images that undergo super resolution.
4) LPIPS: LPIPS calculates the separation between two images in a deep feature space, based on the idea that the perceived similarity of two images is influenced by the way the human visual system processes and interprets them. It measures the deviation of the processed image from the reference image by calculating the distance between the images in a feature space using a deep convolutional neural network. LPIPS score is calculated as the distance between the features of the processed and reference images, and it is more in accordance with human judgment than PSNR and SSIM. The smaller the LPIPS, the more similar the two photos are.
5) PIQE: PIQE, which stands for Perception-based Image Quality Evaluator, is a type of image quality metric that can assess the quality of an image without needing a reference image for comparison. It works by dividing the image into non-overlapping blocks and analyzing each block individually to detect any distortions and assign a quality grade. This approach takes into account the human visual system's tendency to focus on areas with high spatial activity. As a result, PIQE can offer a map with high spatial quality. By combining the quality scores of each block, the overall quality of the image can be determined in PIQE.
6) NRQM [24]: A way to assess the quality of super-resolved images is through the use of a learned no-reference quality metric (NRQM). This NRQM employs three types of statistical data, namely, local frequency features, global frequency features, and spatial features, which are utilized to estimate the perceptual scores of super-resolved images [24]. To achieve this, three regression forests and a linear regression model are trained with a large dataset of super-resolved images that have perceptual ratings. The predicted visual quality of the super-resolved images using the NRQM is in good agreement with the subjective evaluation of the SR results.
7) MOS: Mean Opinion Score (MOS) is a commonly used tool for subjective assessment of the perceived quality of a digital image or video. It is calculated by having a group of human evaluators rate the quality of the super-resolved image on a scale from 1 to 5 and taking the average of their ratings. MOS is considered more reliable and accurate than objective metrics like PSNR or SSIM, but it can be unreliable if there are few human raters. Other common loss functions used in image quality assessment include L1 loss, perceptual loss, SSIM loss, and adversarial loss. To increase the reliability of the MOS, it is generally recommended to use a larger group of

evaluators and a diverse set of reference and super-resolved images. Additionally, it is important to note that MOS can be affected by factors such as personal preferences or bias, and it may not always accurately reflect the perceived quality of the image. Therefore, it is important to use multiple evaluation methods to ensure accurate image quality assessment.

8) FID: FID (Frechet Inception Distance) is a measure utilized for assessing the likeness and quality of images produced by super-resolution algorithms. It compares the generated images to a reference dataset, calculating the distance in feature space. FID differs from PSNR and SSIM in that it measures the dissimilarity between the high-resolution input (IHR) and the generated image output (ISR). It is often employed specifically for assessing the quality of facial images. It uses an Inception network to extract features and calculates the distance between means and covariance matrices. The FID is then calculated as the distance between the means of the feature vectors of the generated and reference images, as well as the distance between their covariance matrices. It can be represented as,

$$FID = \|\mu X - \mu Y\| \tag{6}$$

Smaller FIDs indicate higher visual quality. In the latent space of Inception-V3 , FID offers a symmetric measure of the separation between two picture distributions. Nash et al. recently introduced sFID as a variant of FID that incorporates spatial information rather than the typical pooling features . According to their research, this metric more accurately depicts spatial linkages and rewards image distributions with coherent high-level structure.

5. METHODS AND TECHNIQUES

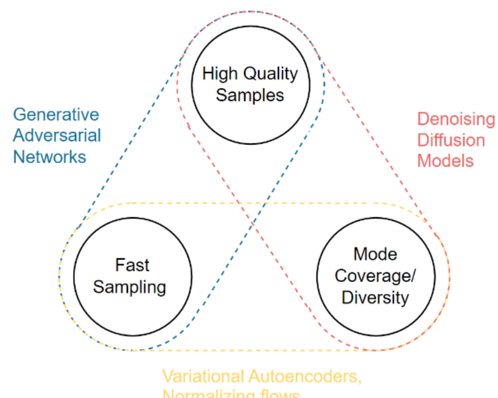

Fig -1: Strengths and Weaknesses of GANs (Generative Adversarial Networks), VAEs (Variational Auto Encoders) and DDMs (Denoising Diffusion Models)

RSISR methods and techniques have been extensively studied by Researchers for the past decade and there have been various approaches that have been developed. Some of such methods and techniques are discussed in the following section. Table-3 gives an overview of the various Image super-resolution techniques. Fig -1: shows the strengths and weaknesses of GANs, VAEs, and DDMs

1) Denoising Diffusion Probabilistic Model: The denoising diffusion probabilistic model (DDPM) is a method for image super-resolution that is based on a diffusion process that is used to denoise and enhance the resolution of images [1], [8], [12]. The DDP model consists of three components: a diffusion process, a probabilistic model, and a noise model. The diffusion process is used to smooth out noise and other distortions in the input image, while the probabilistic model is used to capture the underlying structure of the image. The noise model is used to model the noise present in the input image, and is used to guide the diffusion process in a way that reduces the noise while preserving the underlying structure of the image. Mathematically, the diffusion process is formulated as: “Given a data distribution $x_0 \sim q(x_0)$, the forward noising process q which produces latents x_1 through x_T by adding Gaussian noise at time t with variance $\beta_t \in (0, 1)$” [1].

$$q(x_1, \dots, x_T | x_0) = \prod_{t=1}^{T} q(x_t | x_{t-1}) \tag{7}$$

$$q(x_t|x_{t-1}) = \mathcal{N}(x_t; \sqrt{1-\beta_t x_{t-1}}, \beta_t I) \tag{8}$$

With enough time and good scheduling of β_t the latent x_T is nearly an isotropic Gaussian distribution. By getting the exact reverse distribution $q(x_{(t-1)}|x_t)$. we can sample $x_T \sim N(0, I)$ and run the process in reverse to get a sample from $q(x_0)$ [1].. One limitation of DDP is that it is computationally intensive to implement, as it involves solving a large number of diffusion equations for each image for thousands of time steps. To tackle this problem [1] has proposed improvements to optimize the denoising process. [8],[11], [12] optimize the reverse denoising process by either implementing a GAN or an estimator to estimate the noise at each step so as to provide a probabilistic guess of the time steps that would be required for an input image. The GANs or the estimator are able to accurately generate noise for Time step T_n as it is a gaussian noise distribution in single computation relieving the need for computing for each time step. Another shortcoming of DDP for image super-resolution is that it is based on a simplified model of the image formation process, and may not accurately capture all of the factors that contribute to image degradation. This can result in suboptimal performance on real-world images that are more complex than the model can handle.

2) Degradation based models : Degradation-based models for image super-resolution are a type of method that aims to improve the resolution of images by modeling the process of image degradation. Degradation-based models for image super-resolution work by building a model of the degradation process that takes place between the high-resolution original image and the low-resolution observed image. The model is then used to "reverse" the degradation process, in order to estimate the high-resolution original image from the low resolution observed image. Degradation model requires a dataset of HR-LR image pair to learn the degradation process. The model could be based on a physical model of image degradation, such as a blur kernel or noise model, or it could be a statistical model that captures the statistical properties of the degradation process. Real ESRGAN [4] proposed a second order degradation process that would better depict real world image degradation which includes multiple blur kernels, noise inducing layers and JPEG compression layers. This helped the achieve excellent results in real world applications. Latent Multi-Relation Reasoning (LAREN) is an image super resolution technique that uses two graph-based approaches to improve the quality of the generated high-resolution images. The first approach, called graph-based disentanglement, uses hierarchical multi-relation reasoning to construct a latent space that is better able to capture the underlying structure of the image. The second approach, called graph-based code generation, uses recursive relation reasoning to generate image specific codes that allow prior GANs to produce desirable image details [2]. When considering super-resolution with scaling factors up to 64x LAREN achieves state-of-the-art results. Degradation-based models for image super-resolution offer a flexible and powerful approach to improving the resolution of images, as they can be tailored to specific degradation processes and can be trained using real-world data. However, they can be computationally intensive to implement and may require a large dataset of high-resolution and low-resolution image pairs in order to achieve good results. It is also worth noting that the performance of degradation-based models for image super-resolution can be improved by using advanced optimization techniques and machine learning algorithms. For example, some methods use deep learning techniques, such as convolutional neural networks (CNNs), to learn more powerful models of the degradation process. Others use optimization algorithms, such as gradient descent, to minimize the difference between the high-resolution original image and the low-resolution observed image. Another way to improve the performance of degradation-based models for image super-resolution is to use data augmentation techniques, which involve generating additional training data by applying various transformations to the existing training data. Data augmentation can help to improve the generalization ability of the model and can lead to better performance on a wider range of images.

Table-3: Overview of different Image Super Resolution Techniques

Method	Published	Category	Keyword
SRDiff	2021[8]	Diffusion Probabilistic based	Likelihood, Gaussian Noise, Markov Chain, Latent space Interpolation, Content fusion, Reverse Diffusion, Conditional Noise Predictor.
SR3	2021[11]	Diffusion Probabilistic based	Stochastic iterative denoising, U-Net, Gaussian noise, Noise scheduling.
IDDPM	2021[1]	Diffusion Probabilistic Model	Log-likelihood, Improving sampling speed, Parameterization and Lhybrid objective.

TSRN	2020[7]	Image Pair Based	TextZoom, Paired low – high-resolution Images with variable focal lengths, Sequence Residual Block, Boundary-aware loss, Alignment, Gradient profile loss.
s-LWSR	2020[9]	Image pair-based	Flexibly adjustable pipeline, U-NET, Depthwise-separable convolution, Symmetric connection framework, Information pool.
DRN	2018[17]	Image pair-based	Inverse/Dual LR-HR mapping, Dual Regression.
DeblurGAN	2022[18]	Image pair-based	Motion Deblurring, Skip connections, Discriminator(patchGAN), Enhanced Feature Propagation, Self-made (1050 photos taken with a camera on 66 mobile devices). Includes Sharp, De-focused and motion Blur images.
Real-ESRGAN	2021[4]	Degradation model-based	High-order Degradation (Close Real world degradation estimation), U-NET discriminator, Spectral normalization.
LAREN	2022[2]	Degradation model-based	Gan-Priors, Stochastic noise injection, Hierarchical muti-relational reasoning, Graph based disentanglement.
D_GAN_ESR	2021[5]	Degradation model-based	Image filtering, Double GANs (Generator), adversarial loss, Peak Signal to Noise Ratio Features (PSNR-F).

3) Image-Pair Based: To improve the accuracy of image super-resolution models, it is necessary to have a dataset of LR-HR image pairs. While these pairs can be created by applying degradation models to high-quality images, it can be challenging to derive realistic degradation models from real-world images. One solution is to collect images of the same scene at different resolutions, which can be used to create realistic LR-HR image pairs for training models such as the RSISR model. This approach allows researchers to overcome the challenge of deriving explicit degradation models from real-world images and improve the performance of image super-resolution models. At present there are three ways of collecting LR-HR image pairs: the focal length adjusting-based approach, the hardware binning-based approach, and the beam splitter-based approach [7], [21]. Image super-resolution for certain types of images, such as text images, remote sensing images, and medical images, often requires specialized models due to the unique characteristics of these types of images. For example, the TSRN method developed by [7] is designed specifically for super-resolution of real scene text images. In addition to the challenges common to all real-world images, it is important to consider the particularities of these types of images when designing image super-resolution models. To improve the accuracy of super-resolution for realistic text images in the TextZoom dataset, the TSRN [7] method uses a spatial transform network-based central alignment module and combines it with a gradient prior loss in addition to the standard L2 loss during training. This helps to align the images and enhance the shape boundary of characters, resulting in increased recognition accuracy for real-world text images. Some challenges faced by image-pair based methods are: Convergence issues, Sensitivity to initialization, Limited ability to handle large upscaling factors, Limited ability to handle multiple types of degradation, Overfitting, Lack of diversity in the training data, Limited ability to capture high-frequency information.

6. CONCLUSION

There has been increasing interest in using super-resolution techniques to improve the resolution of real-world images. This paper reviews several approaches that have been used for this purpose which include degradation modeling-based algorithms, image pairs-based algorithms, Denoising diffusion models and various Deep-learning techniques. The paper also discusses common datasets and evaluation metrics used to train and test super-resolution models, and identifies challenges that remain to be addressed, such as the need for more realistic datasets, specific models for real-world image super-resolution, and improved methods for evaluating the performance of these techniques. DDPMs show immense promise in achieving low log-likelihood and mitigates the problem of mode collapse with high fools rates when tested for human evaluation. In spite of these promising results, they still under perform in comparison with state-of-the-art GANs in terms of fidelity and image quality. Further research and improvements in this field can help mitigate problems and challenges faced by both GANs and DDPMs.

6. REFERENCES

[1] Nichol and P. Dhariwal, "Improved Denoising Diffusion Probabilistic Models." arXiv: 2102.09672, 2021.

[2] Jiahui Zhang, Fangneng Zhan, Yingchen Yu, Rongliang Wu, Xiaoqin Zhang, Shijian Lu "Latent Multi-Relation Reasoning for GAN-Prior based Image Super-Resolution," arXiv:2208.02861 [cs.CV],2022.

[3] T. Karras, S. Laine, and T. Aila, "A Style-Based Generator Architecture for Generative Adversarial Networks." arXiv:1812.04948, 2019.

[4] Xintao Wang, Liangbin Xie, Chao Dong, Ying Shan "Real-ESRGAN: Training Real-World Blind Super-Resolution with Pure Synthetic Data," IEEE/CVF International Conference on Computer Vision Workshops (ICCVW), 2021.

[5] A. Hamdi, Y. K. Chan, and V. C. Koo, "A New Image Enhancement and Super Resolution technique for license plate recognition," Heliyon, vol. 7, no. 11, Nov. 2021.

[6] Aakerberg, K. Nasrollahi, and T. B. Moeslund, "Real-World Super-Resolution of Face-Images from Surveillance Cameras." arXiv:2102.03113, 2021.

[7] W. Wang et al., "Scene Text Image Super-Resolution in the Wild." arXiv: 2005.03341, 2020.

[8] H. Li et al., "SRDiff: Single Image Super-Resolution with Diffusion Probabilistic Models." arXiv: 2104.14951, May 18, 2021.

[9] Biao Li, Bo Wang, Jiabin Liu, Zhiquan Qi, and Yong Shi, "s-LWSR: Super Lightweight Super-Resolution Network," IEEE TRANSACTIONS ON IMAGE PROCESSING, VOL. 29, 2020.

[10] Junjun Jiang, Chenyang Wang, and Xianming Liu, "Deep Learning-based Face Super-resolution: A Survey," ACM Comput. Surv. 55, 1, Article 13 2021.

[11] C. Saharia, J. Ho, W. Chan, T. Salimans, D. J. Fleet, and M. Norouzi, "Image Super-Resolution via Iterative Refinement." arXiv: 2104.07636, 2021.

[12] P. Dhariwal and A. Nichol, "Diffusion Models Beat GANs on Image Synthesis." arXiv: 2105.05233, Jun. 01, 2021.

[13] L. Wang et al., "Unsupervised Degradation Representation Learning for Blind Super-Resolution." arXiv: 2104.00416, 2021.

[14] H. Chen et al., "Real-world single image super-resolution: A brief review," Information Fusion, vol. 79, pp. 124–145, Mar. 2022.

[15] Sjoerd van Steenkiste, Karol Kurach, Jürgen Schmidhuber, Sylvain Gelly, "Investigating object compositionality in Generative Adversarial Networks", Neural Networks, Volume 130, 2020.

[16] M. Fritsche, S. Gu, and R. Timofte, "Frequency Separation for Real-World Super-Resolution." arXiv: 1911.07850, 2019.

[17] Karras et. al, "Progressive Growing of GANs for Improved Quality, Stability, and Variation", in Internation Conference on Representation Learning (ICLR), 2018.

[18] Abedjooy, Aref \& Ebrahimi, Mehran. "Multi-Modality Image Super-Resolution using Generative Adversarial Networks," arXiv.2206.09193, 2022.

[19] K. V. Suresh, G. M. Kumar and A. N. Rajagopalan, "Superresolution of License Plates in Real Traffic Videos," in IEEE Transactions on Intelligent Transportation Systems, vol. 8, no. 2, pp. 321-331, June 2007.

[20] Y. Guo et al., "Closed-loop Matters: Dual Regression Networks for Single Image Super-Resolution." arXiv, May 22, 2020.

[21] L. Kuhlane, D. Brown, A. Boby, and M. Marais, "Deep Learning Approach to Image Deblurring and Image Super-Resolution using DeblurGAN and SRGAN," Sep. 2022.

[22] P. Wei, Z. Xie, H. Lu, Z. Zhan, Q. Ye, W. Zuo, and L. Lin, "Component divide-and-conquer for real-world image super-resolution," in European Conference on Computer Vision (ECCV), 2020.

[23] H. R. Sheikh, A. C. Bovik, and G. De Veciana, "An information fidelity criterion for image quality assessment using natural scene statistics,"IEEE Transactions on Image Processing, vol. 14, no. 12, pp. 2117–2128, 2005.

[24] C. Ma, C.-Y. Yang, X. Yang, and M.-H. Yang, "Learning a no-reference quality metric for single-image super-resolution," Computer Vision and Image Understanding, vol. 158, pp. 1–16, 2017.