Learning Degradation Using Generative Adversarial Networks for Image Super-Resolution

Molefe Molefe
University of the Witwatersrand
School of Computer Science and Applied Mathematics
Johannesburg

Richard Klein
University of the Witwatersrand
School of Computer Science and Applied Mathematics
Johannesburg

***Abstract*—Image super-resolution is a research endeavour that has gained notoriety topic in the field of computer vision. The goal is to increase the spatial dimensions of an image while maintaining the perceptual quality. Recent methods employ deep learning methods to reconstruct significant details prevalent in low-resolution images. Deep learning is a computationally expensive research endeavour that requires tedious methods of collecting data. Super-resolution tasks require low-resolution, high-resolution corresponding image pairs to solve the problem. Currently, the problem is solved by creating synthetic low-resolution representations of corresponding high-resolution images. Unfortunately, recent methods employ ineffective downscaling operations to achieve synthetic low-resolution images. These methods fail to generalise well on real-world images with different degradations. We propose a model that solves such a problem by learning the degradation using a generative adversarial network. Our proposed model appropriately super-resolves images with unknown degradation while reconstructing perceptual details from low-resolution inputs. We demonstrate that learning to downsample images in a self-supervised manner is a worthwhile and viable approach for super-resolution.**

***Index Terms*—Super-resolution, Image Reconstruction, Computer Vision**

I. Introduction

Image super-resolution refers to increasing the resolution of an image while improving the perceptual quality. The completion of a super-resolution task requires corresponding low-resolution, high-resolution image pairs, which we refer to as LR-HR pairs, for supervision. Many approaches attempt super-resolution on bicubically downsampled images [1, 2, 3, 4]. Currently, super-resolution techniques fail to attain high-perceptive quality consistency in pixels across the entire generated image. Recent methods fail to produce pleasing results when attempting super-resolution on real-world low-resolution images. We aim to compare whether applying state-of-the-art super-resolution techniques on learned degradation performs better than super-resolution on bicubically downsampled images. Our investigation primarily focuses on modelling low-resolution images that are later super-resolved. We generate synthetic low-resolution images by training a self-supervised generative adversarial network. This focus isolates the problem to realise high-resolution images while considering real-world degradation. However, a real-world application of super-resolution would not provide high-resolution counterparts. A generalisable model, specific on training contextual information, that super-resolves any low-resolution image is desirable. Generative adversarial networks learn the distribution of the dataset they are trained on. As a result, super-resolution generative adversarial networks perform well on test data that have a similar environment to the training data. We investigate the plausibility of eradicating the need for supervision in super-resolution tasks. We aim to train super-resolution models without corresponding image pairs allowing us to apply super-resolution at scale. Reducing supervision means decreasing the cost of collecting large samples of data. Our research effectively expedites the research endeavour of image super-resolution by reducing cost whilst attaining good perceptual quality of super-resolved images. In doing so, we propose a two-stage GAN that combines a High-To-Low GAN to learn degradation [5] and ESRGAN [6] to super-resolve images. We empirically measure the similarity of target high-resolution images and super-resolved images using the *Peak-Signal Noise Ratio* and *Structural Similarity Index*. Our proposed method performs well perceptually and produces reasonable quantitative results.

Fig. 1: Proposed method (right) appears similar to the original image (left) with PNSR of 19.39 dB

II. Related Work

Deep generative models proposed by Goodfellow et al. [7] have displayed an impressive ability to generate content such as images. Current methods have proposed an unsupervised approach to apply super-resolution on low-resolution images and produce results that downscale correctly [8]. Traditional methods that generate artificial low-resolution images prevent

generative models from learning textures present in images and fail to generalise well on real-world examples [5]. Generative models currently present a problem for intelligent image synthesis and image editing. Gu et al. [4] argues that super-resolution methods that partition the image into overlapped patches and process patches separately ignore the consistency of pixels in overlapped pixels.

A. Image Upsampling

Image upsampling refers to increasing the spatial resolution of images. Recent methods focus on modelling real-world low-resolution input images. Given the difficulty of data acquisition, researchers aim to find effective ways to improve the reconstruction quality of super-resolution methods. Supervised super-resolution fails dismally with real-world degradation such as sensor noise and compression artifacts. Zhang et al. [9] investigate such issues and propose frameworks to model real-world degradation. They achieve this by applying various techniques to downgrade the perceptual quality and downsample the spatial resolution. Smaller filter sizes [10] with more mapping layers are argued to increase the speed of super-resolution models that reconstruct the image quality of an image. These methods enable the network to learn upsampling filters directly from the low-resolution representation.

1) Post-Upsampling: Post-upsampling methods perform convolutional operations and feature extraction on low-resolution representations. They also replace predefined upsampling interpolation-based layers with learnable layers embedded in deep learning models. Feature extraction occurs within the low-resolution representation while upsampling occurs later within the model. Researchers [11, 12] believe integrating post-upsampling with deep learning improve computational efficiency. However, scale factors are predetermined before training and thus cannot produce multi-scale super-resolution.

2) Pre-Upsampling: Pre-upsampling methods utilize upsampling algorithms to obtain high-resolution outputs. Deep neural networks refine the resultant feature maps helping models learn an end-to-end mapping from interpolated low-resolution images to high-resolution generated images. These methods introduce artifacts and checkerboard patterns corresponding to image degradation. This occurs as a result of performing convolutional operations and feature extraction on high-resolution representations [3].

B. Image Super-Resolution

Image super-resolution attempts to reconstruct perceptual details from low-resolution images. Deep convolutional networks [3] learn an efficient mapping from low-resolution inputs to high-resolution outputs. Dong et al. [3] mentions that their proposed SRCNN (Super-Resolution Convolutional Neural Network) can learn super-resolution while coping with three-channel colour images. Researchers believe deep learning is necessary for approaching super-resolution at a larger scale.

1) Learning Degradation: Real-world LR images are noisy as a result of degradations and nuisance factors such as sensor noise, compression artifacts, noise and blur. These factors are usually unknown and may appear randomly through the pixel space of an image. Researchers [5, 13] argue that ineffective low-resolution models may lead to poor performance during test time. Recent techniques [2] perform bicubic downsampling with anti-aliasing to downsample the spatial resolution of an HR image. Synthetic LR images generated from predefined convolutional operations fail to capture real-world noise. The performance of super-resolution methods degrade because of their inability to effectively model variations of noise in LR images. Hence, generating LR images via downsampling HR images negatively affects the practical applications of super-resolution [14]. Other works combat such a problem by modelling degradation as a combination of several convolutional operations [15]. Gu et al. [16] argue that combinatoric degradational models are closer to mimicking real-world degradations.

C. Convolutional Neural Networks

Convolutional neural networks are neural networks with convolutional operators that have been widely used to solve computer vision problems. They are particularly useful in finding meaningful relationships between neighbouring pixels present in images. Following the success of convolutional neural networks in computer vision tasks [17], state-of-the-art super-resolution techniques leverage CNNs to reconstruct the image quality of a super-resolved image. Consequently, recent deep learning methods leverage the depth of convolutional neural networks by introducing small convolutional filters with a small receptive field to capture orientation [18]. The consensus surrounding deep learning networks is centred around vanishing or saturating nonlinearities. Ioffe and Szegedy [19] discuss a phenomenon called covariate shift which describes the slow training of a deep network due to low learning rates or specific parameter initialization. Batch normalization circumvents slow training by normalizing layer inputs allowing the models to utilize higher learning rates for faster convergence. However, accuracy saturation and degradation are consequential with the convergence of a normalized deep network. As a result, researchers addressed degradation by introducing deep residual frameworks [20]. Fundamentally, the increasing depth increases the number of weights in a network increasing the likelihood of overfitting. Deeply-recursive convolutional neural networks address overfitting by repeatedly applying the same convolutional layer. Resultant feature maps from recursive convolutional layers reconstruct pixels in high-resolution images. Skip connections [21] store the input signal during recursions yielding contextual information and allowing the copy of the input signal to be used during target prediction. Generated high-resolution images often lack significant details from their corresponding high-resolution target. Ideally, we would hope that super-resolution closely realises the latent details in image quality. Ledig et al. [2] state that certain loss functions are responsible for the semantic information within images, such as texture. Furthermore, they believe that

MSE (mean squared error) averages the pixel loss between the target and a generated high-resolution resulting in overly smooth image patches.

D. Generative Adversarial Networks

The zero-sum game is a game in which one player's gain is another's loss. Games, like chess, characterize such a situation where the net benefit for individual players is eventually zero. A generative adversarial network is an architecture proposed by [7] that suggests two models which compete against each other. The generative model, G, is trained to generate the data distribution. And the discriminator D has the responsibility of determining if a sample belongs to the data distribution generated by G or training data. In other words, the generator attempts to fool the discriminator that the sample belongs in the training data.
Zero-sum games are often associated with the min-max theorem, thus it follows that the adversarial loss function for training generative models is formulated as

$$\min_G \max_D \mathbb{E}_{x \in X}[\log \text{ D(x)}] + \mathbb{E}_{z \in Z}[log(1 - D(G(z)))] \quad (1)$$

where x is sampled from the original data and z is sampled from the generated data.
The generator learns a generator distribution by building a mapping function from a prior noise distribution to data space while the discriminator returns the probability that the input belonged to the training data.

1) Super-Resolution GAN: Ledig et al. [2] paper proposed a *super-resolution GAN* that increases the spatial resolution of an image while improving the perceptual quality. The paper argued that previous loss functions fell short at recognizing subjective qualities in pixels. Hence, the paper introduces a perceptual loss function which consists of *adversarial loss and content loss*. Authors borrowed the notion of residual networks with skip connections [22] to address vanishing gradients in their model.

E. Loss Functions

1) Pixel-Wise Loss: Per-pixel loss measures the pixel-wise difference between output and target images. Additionally, minimising the mean squared error [23] finds the pixel-wise averages of plausible solutions resulting in overly-smooth details in pixels with textual context.

2) Perceptual Loss: Johnson et al. [24] argue that per-pixel loss does not capture perceptual differences between output and target images. Optimizing perceptual loss, based on high-level features extracted from pre-trained networks, generates high-quality images. Specifically, papers define perceptual losses to measure high-level perceptual and semantic differences between output and target images. Style transfer methods utilized style reconstruction loss to penalize differences in style while using the feature reconstruction loss to penalize the output when it deviates from the content in the original target.

F. Image Quality Assessment

Image quality assessment refers to the qualitative and quantitative measurements of super-resolution methods. We consider the objective methods to assess the quality of results.

1) Structural Similarity: The structural similarity index measures(SSIM) the similarity between two images. It is useful when comparing the ground truth image and image generated by super-resolution algorithms. The following equations describe the SSIM as [25]:

$$\text{SSIM}(x, y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)} \quad (2)$$

The equation above is described by the following variables

- $\boldsymbol{\mu_x}$: the average of x
- $\boldsymbol{\mu_y}$: the average of y
- $\boldsymbol{\sigma_x^2}$: the variance of x
- $\boldsymbol{\sigma_y^2}$: the variance of y
- $\boldsymbol{\sigma_{xy}}$: the covariance of x and y
- $\boldsymbol{L}$: the dynamic range of the pixel values
- $\boldsymbol{c_1 = (k_1 L)^2, c_2 = (k_2 L)^2}$: stabilize the division with weak denominator

2) Peak Signal-To-Noise Ratio: A well-used metric used for image quality is the peak signal-to-noise ratio. It is defined as the ratio between the maximum value of a signal and the power of distorting noise that affects the quality of its representation. The following describes its mathematical formulation [25]:

$$PSNR = 10 \cdot \log_{10}\left(\frac{MAX_I^2}{MSE}\right) \quad (3)$$

where MAX_I represents the possible pixel value of the images. The MSE is defined as follows

$$MSE = \frac{1}{mn}\sum_{i=0}^{m-1}\sum_{j=0}^{n-1}[I(i,j) - K(i,j)]^2 \quad (4)$$

where I is an image and K is its noisy approximation.
Ledig et al. [2] set the SRGAN with high upscaling factors up to four measured by the PSNR. The PSNR, measured in decibels, struggles to assess the quality of the image and does not correlate well with human judgement. There is evidence of super-resolution methods that record-high PSNR values corresponding with images that result in overly smooth patches [2].

III. Method

A. Overview

We investigate the effectiveness of learned degradation in the practical application of super-resolution. We approach this goal by using the High-To-Low GAN [5] to learn degradation and super-resolve the result using the ESRGAN architecture [6]. We define the following equations

$$\hat{I}_{LR} = \mathbb{G}_{\downarrow}(I_{HR}; \theta_{LR}) \quad (5)$$

where $G_{\downarrow}$ downsamples a high-resolution image.

$$\hat{I}_{SR} = \mathbb{G}_{\uparrow}(\hat{I}_{LR}; \theta_{HR}) \quad (6)$$

where $G_{\uparrow}$ upsamples the result of the learned degradation.

B. Image Super-Resolution

Super-Resolution describes an image upsampling technique that increases the perceptual quality of a low-resolution image. The trained models perform image-to-image translation by mapping low-resolution images to high-resolution images. Supervised super-resolution depends on the acquisition of LR-HR pairs. We investigate effective techniques to model real-world low-resolution images to avoid taking photographs of LR-HR image pairs. Common interpolation-based downsampling techniques, such as bicubic downsampling, smoothen the texture details in high-resolution images, disallowing super-resolution models from representing original texture details in real-world low-resolution images. Ultimately, the goal is to model a real-world low-resolution input to improve the generalisability of a super-resolution model. Super-resolving bicubically downsampled images presents artifacts and undesirable details in generated HR images. Consequently, we focus specifically on improving the perceptual quality rather than simply increasing the spatial resolution of a low-resolution image. Hence, a common technique is to downsample the spatial resolution followed by quality degradation to generate corresponding LR images from HR images [25].

$$\hat{I}_{LR} = \mathbb{D}(I_{HR}, \theta_{LR}) = (I_{HR} * k) + \eta \tag{7}$$

where $\mathbb{D}$ applies a downsampling technique by applying convolution k on an image. η represents additive noise that may be used to incorporate additional information on low-resolution pixels. Previous methods that approach super-resolution using LR-HR pairs heavily depend on supervision. Obtaining LR-HR pairs is a considerably expensive task. Data collection requires photographing both image pairs under exact circumstances. Ideally, we would prefer general models that can upsample images without corresponding low-resolution representation.

C. Two-Stage Adversarial Architecture

We implemented a two-stage GAN [5] to allow super-resolution methods to learn nuisance factors from real-world degradations. We achieve this by learning degradation from non-corresponding image pairs using a generative model. Our ultimate goal is to reduce the cost of producing image pairs required for training super-resolution images. The *High-To-Low* generative model aims to capture the distribution of different LR representations. We hypothesize that sampling from an LR generative model will be closer to real-world noise than bicubically downsampled images with anti-aliasing.

1) High-To-Low GAN: The *High-To-Low GAN* is tasked to learn how to generate LR images. The generated LR images will serve as inputs to the *Low-To-High* network to train a super-resolution model. Given that LR images are subject to unknown degradations, We borrow from Bulat et al. [5]'s paper an architecture that learns to increase the quality of real-world images corrupted by noise and artifacts. It models a variety of image degradational noise by learning a probability distribution that samples LR images.

2) Low-To-High GAN: The low-to-high GAN is responsible for generating HR images with realistic textures. In particular, the model follows an ESRGAN architecture [6] that improves the perceptual loss by using pre-trained features before activation. Wang et al. [6] argues that their model achieves a more pleasing quality with realistic textures in comparison to Ledig et al's paper.

D. Loss Functions

We aim to effectively super-resolve LR images by using loss functions measuring reconstruction error and guiding the super-resolution models' optimization. Ledig et al's paper [2] proposed a loss function that takes advantage of pre-trained models to infer texture in image details. The ESRGAN further extends the adversarial loss to incorporate texture information in the adversarial training of the generator. Through adequate training, the generator will produce outputs sampled from a distribution learned from adversarial training.

1) High-To-Low GAN: **Pixel Loss**. Pixel loss measures the pixel-wise distance between two images

$$l_{\text{pixel}} = \frac{1}{WH} \sum_{i=1}^{W} \sum_{j=1}^{H} \left(F\left(I^{hr}\right)_{i,j} - G_{\theta_G}\left(I^{d}\right)_{i,j} \right)^2 \tag{8}$$

Hinge Loss. Generative adversarial networks are notorious for training instability, often caused by random noise generation, mode collapse, and vanishing gradient. Miyato et al's paper [26] proposed spectral normalization to tackle such problems presented by adversarial training using GANs.

$$l_{GAN} = \mathop{\mathbb{E}}_{x \sim \mathbb{P}_r} [\min(0, -1 + D(x))] + \mathop{\mathbb{E}}_{\hat{x} \sim \mathbb{P}_g} [\min(0, -1 - D(\hat{x}))] \tag{9}$$

where $\mathbb{P}_r$ is the data distribution and $\mathbb{P}_g$ is the generator G distribution defined by $\hat{x} = G(x)$. **High-To-Low Loss**. Bulat et al. [5] argue that a GAN-centered approach motivates the GAN to drive the image generation process. As a result, L_2

pixel loss is responsible for refining LR images to make them sharper for super-resolution.

$$l = \alpha l_{\text{pixel}} + \beta l_{GAN} \tag{10}$$

where α and β are the corresponding weights and $\beta l_{GAN} > \alpha l_{\text{pixel}}$ in general.

2) Low-To-High GAN: **Perceptual Loss**. The perceptual loss constrains the pre-trained model features before activation to model perceptual similarity between real HR images and super-resolved images.

$$L_G = L_{\text{percep}} + \lambda L_G^{Ra} + \eta L_1 \tag{11}$$

where $L_1 = \mathbb{E}_{x_i} \|G(x_i) - y\|_1$ is the content loss that evaluates distance between generated super-resolved image and real HR image. Wang et al's paper [6] argues that perceptual loss will prevent the drawbacks presented in Ledig et al. [2]'s SRGAN paper.

Adversarial Loss. The ESRGAN employs a relativistic GAN [27] to predict the probability that real images are relatively more realistic than images generated from the fake dataset. Researchers argue relativistic GANs from a super-resolution perspective recover more detailed textures. The generator adversarial loss is defined as:

$$L_G^{R_a} = -\mathbb{E}_{x_r}[\log(1 - D_{R_a}(x_r, x_f))] - \mathbb{E}_{x_f}[\log(D_{R_a}(x_f, x_r))] \tag{12}$$

where $x_f = G(x_i)$ and x_i stands for the input LR image. It is observed that the adversarial loss for generator contains both x_r and x_f. The discriminator adversarial loss is defined as:

$$L_D^{R_a} = -\mathbb{E}_{x_r}[\log(D_{R_a}(x_r, x_f))] - \mathbb{E}_{x_f}[\log(1 - D_{R_a}(x_f, x_r))] \tag{13}$$

IV. Experiments

A. Dataset

We trained our two-stage GAN on the DIVerse 2K resolution image dataset taken from the NTIRE 2017 challenge on *single image super-resolution* [28]. The data was largely collected from the internet and contains 1000 images with very high resolution. Images are degraded via two methods:

- **Standard:** Bicubic downsampling with anti-aliasing
- **Unknown Downscaling Operations:** Diverse set of dowsampling techniques with additive noise to model real-world degradation.

B. Train-Validation-Test Split

- **Train Data:** 800 High-Resolution images
- **Validation Data:** 100 High-Resolution Images
- **Test Data:** 100 High-Resolution Images

We trained 800 images from the DIVerse 2K dataset to train the two-stage GAN to super-resolve images. The validation dataset aided us in fine-tuning the model by selecting the best hyperparameters. We use the test to evaluate and report the quality of the model using image quality assessment techniques.

C. Data Preparation

Image Size. High-Resolution images have relatively high perceptual detail and size. Such images present a problem for training expansive models. Hence, we only consider random crops of size 256 $\times$ 256 for each LR-HR pair from the dataset to alleviate pressure on the GPU.

D. Training

1) Overview of Training Process: The *High-To-Low GAN* is responsible for generating synthetic low-resolution images. We collect high-resolution images that have corresponding low-resolution counterparts with unknown degradations. In doing so, we randomize the order of the low-resolution images to train our *High-To-Low GAN* in an unsupervised manner. Thus, this model utilizes self-supervision to generate low-resolution images. The result is directly fed into the ESRGAN architecture to generate corresponding high-resolution images. We train both models with different hyperparameters to achieve good results. The self-supervised training allows our method to perform well independent of corresponding image pairs.

2) Hyperparameters: The hyperparameters chosen for the *High-To-Low GAN* are listed as follows [5]:

- **Learning Rate:** $1e^{-4}$
- **Epochs:** 1409
- **High-To-Low Loss:** $\alpha = 1, \beta = 0.05$, (10)
- **Adam Optimization:** $\beta_1 = 0, \beta_2 = 0.9$

We train a generative model (ResNet) before training the GAN. The resultant ResNet model is further fine-tuned using an adversarial approach to learn texture. The hyperparameters chosen for the ESRGAN architecture are listed as follows [6]:

- **Learning Rate:** $1e^{-4}$
- **Epochs:** 1409
- **Perceptual Loss:** $\lambda = 5 \times 10^{-3}, \eta = 1 \times 10^{-2}$ (11)
- **ADAM Optimization:** $\beta_1 = 0.9, \beta_2 = 0.999$

E. Results

Fig. 2: Top-left (Ground Truth), Top-right (Enhanced SR-ResNet), Bottom-left (Enhanced SRGAN), Bottom-right (Proposed Two-Staged ESRGAN)

TABLE I: Comparison of SRGAN [2], ESRGAN [6], and our proposed two-stage GAN. The table stores the PNSR (with standard deviation) values measured in decibels and the SSIM (with standard deviation) that measures the similarity between the produced output and target image

Dataset	SRGAN	ESRRestNet	ESRGAN	ESRGAN w/o Gradient Penalty	Two-Stage ESRGAN (Ours)
Set5 [29]					
PSNR (dB)	19.86 (1.28)	25.15 (2.67)	24.40 (1.40)	**27.83** (2.51)	17.27 (2.48)
SSIM	0.59 (0.10)	0.70 (0.15)	0.75 (0.11)	**0.80** (0.10)	0.62 (0.07)
Set14 [30]					
PSNR (dB)	20.11 (1.91)	**27.61** (2.49)	22.89 (2.48)	25.72 (2.88)	17.23 (1.88)
SSIM	0.55 (0.13)	**0.80** (0.09)	0.66 (0.16)	0.70 (0.15)	0.57 (0.12)
BSD100 [31]					
PSNR (dB)	19.84 (2.14)	23.06 (2.96)	22.14 (2.91)	**23.24** (2.96)	17.30 (2.77)
SSIM	0.53 (0.10)	0.66 (0.13)	0.64 (0.13)	**0.66** (0.13)	0.55 (0.13)

The ESRGAN, in Fig. 2,performs relatively well regarding perceptual quality in comparison to the other methods. Also, the ESRGAN trained without gradient penalty outperforms the rest of the methods in terms of PSNR and SSIM (refer to Table I). We generated the results by running each of the models on the individual test datasets [29, 30, 31] and averaged the image quality assessment recordings. Our proposed two-staged ESRGAN, taken from Fig. 2, trains a model that learns a low-resolution representation that is super-resolved later in the ESRGAN generator. The model produces outputs with a comparatively higher dynamic range than the rest. Also, the lighter and darker regions of the image appear to be emphasized. Resultant images demonstrate a colour space that is more luminant and bright in comparison to the other methods. Although their structural similarity seems relatively low, these results may be desirable in a real-world context. Generally, saught models are those that generalise well on unseen data. Consider LR datasets without corresponding HR counterparts. Images with a broader colour space may be more desirable in terms of perceptual quality.

F. Limitations and Failures

1) High-To-Low GAN: In theory, generating LR images that capture variations that resemble real-world noise is plausible. However, generated HR images produce noise in dark regions of an image. The image dynamic range is stretched farther resulting in overwhelming dark patches in regions where the original image was dark. After investigating the result of the *High-To-Low GAN*, we hypothesize that since LR images have smaller spatial dimensions, generated LR images are prone to artifacts and pixels that erroneously summarize contextual information. Implying that noise present in LR images will also be super-resolved thus magnifying the presence of noise in generated HR images.

2) Misrepresenting Colour: Our method emphasizes specific colours on patches that have densely packed similar pixels. It shows that the model clumsily super-resolves low-resolution patches that have an affinity for dark neighbourhoods. Fig. 3 demonstrates that the model assumed that all patches that had low lighting had a purple colour. This is due to the *High-To-Low* GAN learning from lacking information. As a result, *Low-To-High* GAN learns a noisy distribution and super-resolves patches that are surrounded by significantly dark pixels.

Fig. 3: Demonstrating the impact of generative adversarial networks on low-spatial information.

V. Future Work and Discussion

A. Super-resolution as an ill-posed problem

Menon et al. [8]'s paper showed that low-resolution images correspond to multiple high-resolution images. Super-resolution is fundamentally ill-posed as a mapping from a low-resolution input results in multiple possible outputs. Each high-resolution output could correctly represent the reconstruction of a low-resolution image. Hence, subjectivity thus becomes a determining factor for perceptual quality. Such idiosyncrasies allow for vast interpretations that are contextual for each scenario.

B. Low Resolution Generative Adversarial Networks

Reconstructing texture and high-quality details in images have gained notoriety in computer vision research. Recent methods have proposed deep and dense networks to understand the relationships between pixels in images. They have also considered optimization techniques to help deep networks converge faster. As a result, learning feature extraction from low-resolution representations has been neglected. We hypothesize that learning how to reduce noise in low-resolution inputs should teach models to achieve pleasing results for

super-resolution using learned degradation. Im et al. [32] proposed a variational autoencoder with a training criterion that learns flexible posterior distributions. They observed that their method can help improve the performance of variational autoencoders. This method can be applied to denoise the degradation in low-resolution images. Thus, allowing super-resolution models to produce results with pleasing textures.

C. Practical Applications of Super-resolution

1) Surveillance: Surveillance technology aims to monitor and observe continuous and ongoing activities. Practical applications of surveillance, such as security, require acquiring high-quality photographic information to make meaningful decisions. These methods often require large and expensive equipment to photograph incidents. Image super-resolution can be used to magnify the noisy details in recorded footage [33].

2) Medical Imaging: Medical institutions capture medical images using a medium that engages an elementary particle (or wave) with the tissue. Technologies that capture the medical photograph suffer from signal limitations such as attenuation and scattering. Spontaneous tissue rhythms, such as a heartbeat or inhalation, may misrepresent the resolution of a medical image. Hence, super-resolution methods may enhance the perceivability of medical images such as ultrasounds [34].

VI. Conclusion

We have demonstrated the extent of applying learned degradation as a method to learn super-resolution. Although ESRGAN performs better while using corresponding image pairs, our proposed approach achieves reasonable metric scores without corresponding image pairs demonstrating the potential of investigating self-supervised super-resolution techniques. Furthermore, we demonstrated the potential of using self-supervised techniques to train super-resolution models. Our proposed method greatly reduces the cost of acquiring supervised image pairs. Additionally, we have highlighted that super-resolved images from learned degradation suffer from low-resolution noise. A promising endeavour for approaching super-resolution using learned degradation should consider employing dense networks to extract features from low-resolution images more effectively.

References

[1] J. Yu, Y. Fan, J. Yang, N. Xu, Z. Wang, X. Wang, and T. S. Huang, "Wide activation for efficient and accurate image super-resolution," *CoRR*, vol. abs/1808.08718, 2018. [Online]. Available: http://arxiv.org/abs/1808.08718

[2] C. Ledig, L. Theis, F. Huszar, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, and W. Shi, "Photo-realistic single image super-resolution using a generative adversarial network," 2017.

[3] C. Dong, C. C. Loy, K. He, and X. Tang, "Image super-resolution using deep convolutional networks," 2015.

[4] S. Gu, W. Zuo, Q. Xie, D. Meng, X. Feng, and L. Zhang, "Convolutional sparse coding for image super-resolution," in *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, December 2015.

[5] A. Bulat, J. Yang, and G. Tzimiropoulos, "To learn image super-resolution, use a GAN to learn how to do image degradation first," *CoRR*, vol. abs/1807.11458, 2018. [Online]. Available: http://arxiv.org/abs/1807.11458

[6] X. Wang, K. Yu, S. Wu, J. Gu, Y. Liu, C. Dong, C. C. Loy, Y. Qiao, and X. Tang, "ESRGAN: enhanced super-resolution generative adversarial networks," *CoRR*, vol. abs/1809.00219, 2018. [Online]. Available: http://arxiv.org/abs/1809.00219

[7] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial networks," 2014.

[8] S. Menon, A. Damian, S. Hu, N. Ravi, and C. Rudin, "Pulse: Self-supervised photo upsampling via latent space exploration of generative models," 2020.

[9] K. Zhang, W. Zuo, and L. Zhang, "Learning a single convolutional super-resolution network for multiple degradations," 2018.

[10] C. Dong, C. C. Loy, and X. Tang, "Accelerating the super-resolution convolutional neural network," 2016.

[11] W. Shi, J. Caballero, F. Huszár, J. Totz, A. P. Aitken, R. Bishop, D. Rueckert, and Z. Wang, "Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network," *CoRR*, vol. abs/1609.05158, 2016. [Online]. Available: http://arxiv.org/abs/1609.05158

[12] C. Dong, C. C. Loy, and X. Tang, "Accelerating the super-resolution convolutional neural network," *CoRR*, vol. abs/1608.00367, 2016. [Online]. Available: http://arxiv.org/abs/1608.00367

[13] S. Zhang, G. Fu, H. Wang, and Y. Zhao, "Degradation learning for unsupervised hyperspectral image super-resolution based on generative adversarial network," *Signal, Image and Video Processing*, vol. 15, pp. 1–9, 11 2021.

[14] H. Chen, X. He, L. Qing, Y. Wu, C. Ren, R. E. Sheriff, and C. Zhu, "Real-world single image super-resolution: A brief review," *Information Fusion*, vol. 79, pp. 124–145, 2022.

[15] J. Gu, H. Lu, W. Zuo, and C. Dong, "Blind super-resolution with iterative kernel correction," *CoRR*, vol. abs/1904.03377, 2019. [Online]. Available: http://arxiv.org/abs/1904.03377

[16] ——, "Blind super-resolution with iterative kernel correction," 2019.

[17] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems 25: 26th Annual Conference on Neural Information Processing Systems 2012. Proceedings of a meeting held December 3-6, 2012, Lake Tahoe, Nevada, United States*, P. L. Bartlett, F. C. N. Pereira, C. J. C. Burges, L. Bottou,

and K. Q. Weinberger, Eds., 2012, pp. 1106–1114.
[18] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2015.
[19] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," 2015.
[20] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," 2015.
[21] J. Kim, J. K. Lee, and K. M. Lee, "Deeply-recursive convolutional network for image super-resolution," 2016.
[22] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," *CoRR*, vol. abs/1512.03385, 2015. [Online]. Available: http://arxiv.org/abs/1512.03385
[23] M. Mathieu, C. Couprie, and Y. LeCun, "Deep multi-scale video prediction beyond mean square error," 2016.
[24] J. Johnson, A. Alahi, and L. Fei-Fei, "Perceptual losses for real-time style transfer and super-resolution," 2016.
[25] Z. Wang, J. Chen, and S. C. H. Hoi, "Deep learning for image super-resolution: A survey," 2020.
[26] T. Miyato, T. Kataoka, M. Koyama, and Y. Yoshida, "Spectral normalization for generative adversarial networks," *CoRR*, vol. abs/1802.05957, 2018. [Online]. Available: http://arxiv.org/abs/1802.05957
[27] A. Jolicoeur-Martineau, "The relativistic discriminator: a key element missing from standard gan," 2018.
[28] E. Agustsson and R. Timofte, "Ntire 2017 challenge on single image super-resolution: Dataset and study," in *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2017, pp. 1122–1131.
[29] M. Bevilacqua, A. Roumy, C. Guillemot, and M. L. Alberi-Morel, "Low-complexity single-image super-resolution based on nonnegative neighbor embedding," 2012.
[30] R. Zeyde, M. Elad, and M. Protter, "On single image scale-up using sparse-representations," in *International conference on curves and surfaces*. Springer, 2010, pp. 711–730.
[31] D. Martin, C. Fowlkes, D. Tal, and J. Malik, "A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics," in *Proceedings Eighth IEEE International Conference on Computer Vision. ICCV 2001*, vol. 2. IEEE, 2001, pp. 416–423.
[32] D. J. Im, S. Ahn, R. Memisevic, and Y. Bengio, "Denoising criterion for variational auto-encoding framework," 2016.
[33] L. Yue, H. Shen, J. Li, Q. Yuan, H. Zhang, and L. Zhang, "Image super-resolution: The techniques, applications, and future," *Signal Processing*, vol. 128, pp. 389–408, 2016. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0165168416300536
[34] D. Kouame and M. Ploquin, "Super-resolution in medical imaging : An illustrative approach through ultrasound," in *2009 IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, 2009, pp. 249–252.