Efficient Single Image Super-Resolution for Images with Spatially Varying Degradations

Simone Bianco(✉), Luca Cogo, Gianmarco Corti, and Raimondo Schettini

Department of Informatics, Systems and Communication, University of Milano-Bicocca, 20126 Milan, Italy
{simone.bianco,luca.cogo,gianmarco.corti,raimondo.schettini}@unimib.it

Abstract. Super-Resolution (SISR) is a computer vision task that aims to generate high-resolution images from their low-resolution counterparts. Typically, Super-Resolution methods use scaling factors of x2, x3, or x4 to uniformly enhance the resolution of the entire image. However, some acquisition devices, such as 360° cameras, produce images with non-uniform resolution across the frame. In this work, we propose to adapt state-of-the-art efficient methods for Single Image Super-Resolution to address the challenge of restoring images affected by spatially varying degradations. Specifically, we focus on the method that won the recent NTIRE 2024 Efficient Super-Resolution Challenge. For our experiments, synthetic images with different spatially varying types of degradation are generated, and the SISR method is specifically modified and trained to effectively handle such challenging scenarios. In addition to evaluating the developed method with traditional image quality metrics such as PSNR and SSIM, we also assess its practical impact on a downstream object detection task. The results on the WIDER FACE face-detection dataset, using the YOLOv8 object detection model, show that applying the proposed SISR approach to images with spatially varying degradations produces artifact-free outputs and enables object detectors to achieve superior performance compared to their application on degraded images.

Keywords: Single image super-resolution · Spatially varying degradation · Efficient · Deep learning

1 Introduction

Single Image Super-Resolution (SISR) is a computer vision task aimed at enhancing the resolution of low-quality images. It has various applications, including medical imaging, satellite image processing, and surveillance [31]. With the advent of deep learning, various methods based on Deep Neural Networks have been proposed and have surpassed traditional methods in various research areas.

E. Rodolà et al. (Eds.): ICIAP 2025, LNCS 16167, pp. 234–246, 2026.
https://doi.org/10.1007/978-3-032-10185-3_19

As a result, even in the research area of Super-Resolution (SR), new methods based on neural networks have been proposed, which are able to obtain better results than traditional techniques [29]. However, despite the progress made by deep learning models, many algorithms require considerable computational and memory capacity, making them difficult to adopt on a large scale or in contexts where resources are limited, as in the case of mobile devices and real-time applications. To this end, the NTIRE Efficient Super-Resolution Challenge [23] was organized, in which participants develop Super-Resolution methods that are as efficient as possible, where efficiency is evaluated in terms of metrics such as runtime, number of parameters, and computational complexity measured in FLOPs. These methods increase the resolution of their input images equally across the entire image; however, there are acquisition devices that generate images that do not have constant resolution. Certain sensors generate images that have spatially variable resolution, as in the case of 360° cameras, cameras with fish-eye lenses, and even acoustic imaging devices. In these scenarios, it can be useful to obtain images with constant resolution without losing information about the details of the content. This topic is still poorly addressed in the literature; in fact, when super-resolution is described, the focus is on the ability of the methods to increase resolution and image quality only in cases where degradation is constant, and these methods are not designed to work well in situations in which images have uneven resolution. The aim of this work is to evaluate the ability of Efficient Super-Resolution (ESR) methods to enhance images with spatially varying resolution. In order to do so, the winner of the NTIRE 2024 Efficient Super-Resolution challenge [23], SPAN [27], has been modified, trained and tested for this purpose. The model has been trained for such situations and tested on various types of degradation to observe whether it could improve the quality of the input images. Lastly, to test the impact of the application of these methods, SPAN was used in combination with YOLOv8 [9] on a face detection dataset to which spatially varying degradations were applied. Various tests were conducted to assess the YOLO model's ability on the face detection task when images present spatially varying resolution.

2 Related Works

Efficient Single Image Super-Resolution. Thanks to the recent rapid development of deep learning techniques, various Single Image Super-Resolution (SISR) methods based on neural networks have been proposed and have outperformed traditional approaches. SRCNN [4] is one of the first deep learning methods applied to the Super-Resolution task, and it is used to learn the mapping between a low-resolution image and its high-resolution version. A disadvantage of SRCNN is that it is computationally inefficient as it works in the space of high-resolution images, since the input image is resized before being processed by the network. In order to address this problem, ESPCN [26] proposed the use of a subpixel convolutional layer called PixelShuffle, which has been used by various methods to increase image resolution only at the end of the method. Other

methods used a different approach, such as SRGAN [13], which is one of the first and most influential applications of the Generative Adversarial Networks [7] for SISR. Recently, SwinIR [14] has been proposed, which is based on Swin Transformers [16] and achieves state-of-the-art results while reducing the number of parameters. Despite the improvements, many methods still require intensive computations or many resources, which may hinder the ability to implement these models even on low-resource devices such as mobile devices. To this end, the NTIRE Efficient Super-Resolution Challenge [23] was organized with the aim of designing methods that are as efficient as possible while still maintaining good quality results. During the various editions of this challenge, many methods have been proposed that are capable of achieving impressive results despite being relatively small and without requiring many resources [10,15,27,30]. To do so, the various teams that participated in the challenge have employed various techniques such as reparameterization, knowledge distillation, and network pruning.

Spatially Varying Blur. Super-resolution can be conceptualized as a two-step process: image upsampling to increase spatial resolution, followed by deblurring to enhance details and recover high-frequency information. While modern approaches integrate these steps into a unified process, where both resolution enhancement and detail recovery are jointly optimized, image upsampling is a simple task and most of the complexity lies in the image deblurring task. Some approaches in the state of the art [5,6] address the combined problem of Super-Resolution and deblurring by explicitly breaking it down into the two stages previously described, i.e. upsampling and deblurring.

Image deblurring is a classic, well-studied computer vision problem. CNN-based models excel at learning data-driven representations and can adapt to a wide range of blur types without explicitly modeling the blur kernel. These methods have shown remarkable performance improvements over traditional techniques, particularly in handling real-world blurs [33]. However, most of the methods assume that the blur is uniform across the entire image. This task becomes more complex when the image presents a spatially varying blur. Several methods have been developed to address these challenges leveraging advancements in computational algorithms, optimization and deep learning. An early approach [20] approximates the blur contained in various regions with a spatially varying kernel using a combination of spatially invariant ones. Then the deblurring is applied to the single regions. Early deep learning deblurring methods were then implemented to remove uniform blur from images such as the method proposed by Schuler et al. [25] which is divided into two stages: the first estimates the blur kernel and the latent image from the blurry image, the second utilizes the blurry image and the latent one to predict the blur kernel. However, real-world images may contain spatially variable blur. In order to address this type of degradation, various methods have been proposed, especially to restore blur caused by moving objects [11,19,24]. Aljadaany et al. [2] proposed DrNet which does not require any estimate of the blur kernel and is further able to invert the effects of the blurring in blind image recovery tasks allowing it to obtain state-of-the-art results on the GoPro dataset [21] and results close to the state of the art

on other datasets while maintaining a low inference time. Others address the problem of spatially varying blur in SR images, such as DMBSR [12], which is capable of achieving better results than other traditional SR methods on images with non-uniform blur.

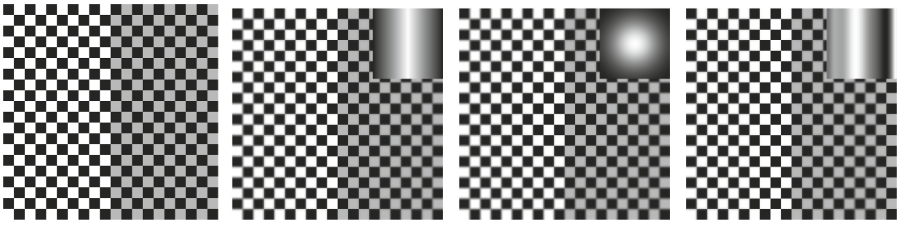
3 Proposed Method

3.1 Degraded Images

Gathering real-world images that are characterized by spatially-varying resolution with their respective counterpart with uniform resolution is a very complex task, since it requires a pair of perfectly geometrically aligned images with two different resolutions in terms of spatial level of detail. For this reason, a synthetically generated dataset is employed in this work, in order to facilitate the experiments discussed in the following sections. To have more control over the type of degradation employed, and the images obtained, it is necessary to generate synthetically degraded images with higher resolution in certain parts of the image and lower in others. Given that super-resolution can be conceptualized as a two-step process, namely image upsampling and image deblurring, with the second one being the most critical one, we focus solely on the latter. Three different types of spatially varying degradation are here used: linear, radial, and random. To simulate a change in resolution, the Gaussian filter and a series of masks are used to make a gradual transition from the original image to the image with the filter applied. The masks used are images with values in the [0,1] range, such that, when applied to the original image, if the values are 1, the output will be the same as the original, while if the values are 0 the output will be the image with the applied Gaussian filter. This results in a blur-like effect that varies along the dimensions of the image itself. Let I_{orig} be the original image, $mask$ the mask used with values between [0,1] and I_{blur} the image on which the Gaussian filter was applied, then the degraded image I_{deg} is obtained as:

$$I_{deg} = I_{orig} \odot mask + I_{blur} \odot (1 - mask) \tag{1}$$

where $\odot$ denotes the element-wise multiplication. By doing so, it is possible to generate images that have a variable blur-like effect by simulating a change in resolution. Figure 1 shows an example of each degradation applied to a checkerboard image with the respective mask applied on it. In these cases, it is possible to observe the different spatially varying blur effects obtained using different degradations. In both linear and radial degradations the loss in detail is constant from the center of the image, where it is present a higher resolution, to the sides where a stronger blur is applied. The difference between the two cases is that in the linear case the transition occurs only along the horizontal axis, while in the radial degradation this takes place also along the vertical axis. Likewise to the first case, with the random degradation, the transition occurs only along the width of the image; however, it is not constant but random. The random mask applied to each image varies for each randomly degraded image. For this



(a) Ground truth (b) Linear (c) Radial (d) Random

Fig. 1. Images obtained by applying the different degradations. The degraded images are obtained by applying the Eq.(1) with the respective mask on the ground truth image. The masks are those in the top right corner of each degraded image, and a pixel value closer to white means a pixel closer to the original.

purpose, random value distributions are used to generate masks, allowing the simulation of random transitions from the original images to the images with the Gaussian filter applied.

3.2 Method Changes

During the experiments described in the next sections, two versions of SPAN [27] have been used. The difference between the two is the number of feature channels employed and therefore the number of parameters. For convenience, the smaller model and winner of the challenge is denoted SPAN-s, while the bigger one is referred to as SPAN. In the first case, the number of feature channels employed by the convolutional layers is 28 and in the second is 48. In both cases, the models were initially trained for 4× upsampling, with a 4× PixelShuffle as the final layer. However, since the image dimensions in our case remain unchanged, the models have been adapted to generate outputs with the same resolution as the input (i.e. the PixelShuffle layer was removed) while enhancing the level of detail. Then the models have been trained to handle the new spatially varying degradations. Both the versions have been trained with the same training configuration, that is: batch size of 16, initial learning rate of $1e^{-4}$ that is reduced every 10,000 iterations by a factor of 0.1, $L1$ loss function, Adam optimizer and 20,000 as the number of training iterations. The working environment employed during the following tests utilizes a Nvidia RTX 4070S GPU.

4 Experimental Results

4.1 ESR Methods for Spatially Varying Resolution

To assess the ability of Efficient Super-Resolution (ESR) models in handling images with spatially varying resolution, the SPAN models, modified as described in Sect. 3.2, have been trained from scratch on degraded images. The training was conducted using three versions of the DIV2K dataset [1], each modified to

Table 1. PSNR and SSIM obtained by applying the methods to the images in the test set. The best values contained in each column are listed in bold.

Model	BSD100					
	Linear		Radial		Random	
	PSNR↑	SSIM↑	PSNR↑	SSIM↑	PSNR↑	SSIM↑
Degraded	30.1696	0.9078	27.2973	0.8140	31.4426	0.9137
SPAN	30.9011	0.9210	**34.9025**	**0.9608**	32.4254	0.9306
SPAN-s	**31.3135**	**0.9310**	31.8414	0.9287	**32.5441**	**0.9409**

Model	Manga109					
	Linear		Radial		Random	
noentry	PSNR↑	SSIM↑	PSNR↑	SSIM↑	PSNR↑	SSIM↑
Degraded	28.8745	0.9450	25.7212	0.8878	29.2848	0.9368
SPAN	30.2921	0.9582	**35.4847**	**0.9797**	**30.9768**	**0.9548**
SPAN-s	**30.4650**	**0.9604**	31.5764	0.9633	30.8488	0.9547

Model	Set5					
	Linear		Radial		Random	
	PSNR↑	SSIM↑	PSNR↑	SSIM↑	PSNR↑	SSIM↑
Degraded	31.4673	0.9488	29.1126	0.9081	30.2307	0.9093
SPAN	32.6783	0.9588	**37.2752**	**0.9801**	31.6903	0.9276
SPAN-s	**33.2132**	**0.9638**	34.5581	0.9684	**32.5492**	**0.9457**

Model	Set14					
	Linear		Radial		Random	
	PSNR↑	SSIM↑	PSNR↑	SSIM↑	PSNR↑	SSIM↑
Degraded	29.9072	0.9192	27.5581	0.8575	31.9502	0.9425
SPAN	30.4842	0.9276	**35.2481**	**0.9691**	**32.6576**	0.9516
SPAN-s	**31.1062**	**0.9402**	32.4917	0.9540	32.4463	**0.9568**

Model	Urban100					
	Linear		Radial		Random	
	PSNR↑	SSIM↑	PSNR↑	SSIM↑	PSNR↑	SSIM↑
Degraded	22.6222	0.7307	21.9710	0.6920	22.4069	0.7141
SPAN	**25.4023**	**0.9381**	**24.8156**	**0.8189**	**25.6190**	**0.8388**
SPAN-s	24.6988	0.8159	24.5486	0.8114	24.4617	0.8052

incorporate a specific degradation: linear and radial degradations were applied using the same mask across all images, while for random degradation, a unique random mask was generated for each image. Similarly, for testing, three degraded versions were created for each of the BSD100 [17], Manga109 [18], Set5 [3], Set14 [32], and Urban100 [8] datasets.

Table 1 reports the PSNR and SSIM values, computed by comparing the degraded images and the outputs of the different methods against the original images. The results indicate that the models are able to effectively enhance image quality, making them more similar to the original counterparts and yielding

Table 2. PSNR value obtained by applying the SPAN-s model trained to handle the degradations present in the first column, on images degraded with the degradation present in the first row. The *Average* column contains the mean value obtained for each row.

Test				
Train	Linear	Radial	Random	Average
Linear	35.5975	30.2463	35.6630	33.8197
Radial	31.9280	36.5738	31.6145	33.5968
Random	35.4167	30.4630	35.5504	33.8725

higher PSNR and SSIM values than the degraded images. Moreover, despite the differences in model parameters, both versions are able to successfully improve degraded images achieving comparable PSNR and SSIM scores, with the only exception being SPAN-s trained on radially degraded images, which exhibits lower PSNR values compared to its larger counterpart.

Further experiments were conducted to assess the SPAN-s model's generalization ability on degradations it was not specifically trained for, with the corresponding PSNR values reported in Table 2. As expected, the highest PSNR values appear along the diagonal, corresponding to models tested on the same degradation type they were trained for. However, when applied to different degradation types, the models yield varying results. Notably, models trained on linear degradation perform well not only on their own degradation but also on random degradation, and vice versa. This is likely due to the fact that, unlike radial degradation, both linear and random cases involve resolution changes only along the image width. The radial case, on the other hand, exhibits uneven resolution even along the height of the image. Despite this, it still achieves better results on other degradation types when compared with the PSNR obtained by the other two methods in the radial degradation case. On average the models obtain similar results despite the differences in terms of degradations.

Figure 2 illustrates visual results for an image extracted from the test set. Notably, along the edges in both the linear and radial cases, certain details remain partially unrecovered. Therefore, even though the models are able to restore many of the details contained in the images, they still struggle in the areas where the degradation is stronger.

4.2 Efficient Super-Resolution for Face Detection

In order to evaluate the effectiveness of applying an Efficient Super-Resolution (ESR) method to images with spatially varying resolution, the previous discussed models were integrated into a face detection pipeline as preprocessing step. A subset of the WIDER Face dataset [28] was used for both training and evaluation, applying the degradations described in Sect. 3.1. The dataset was divided into 1,000 images for training, 200 for validation, and 500 for testing. Three different versions of YOLOv8 [9], with different sizes, were used as object detection

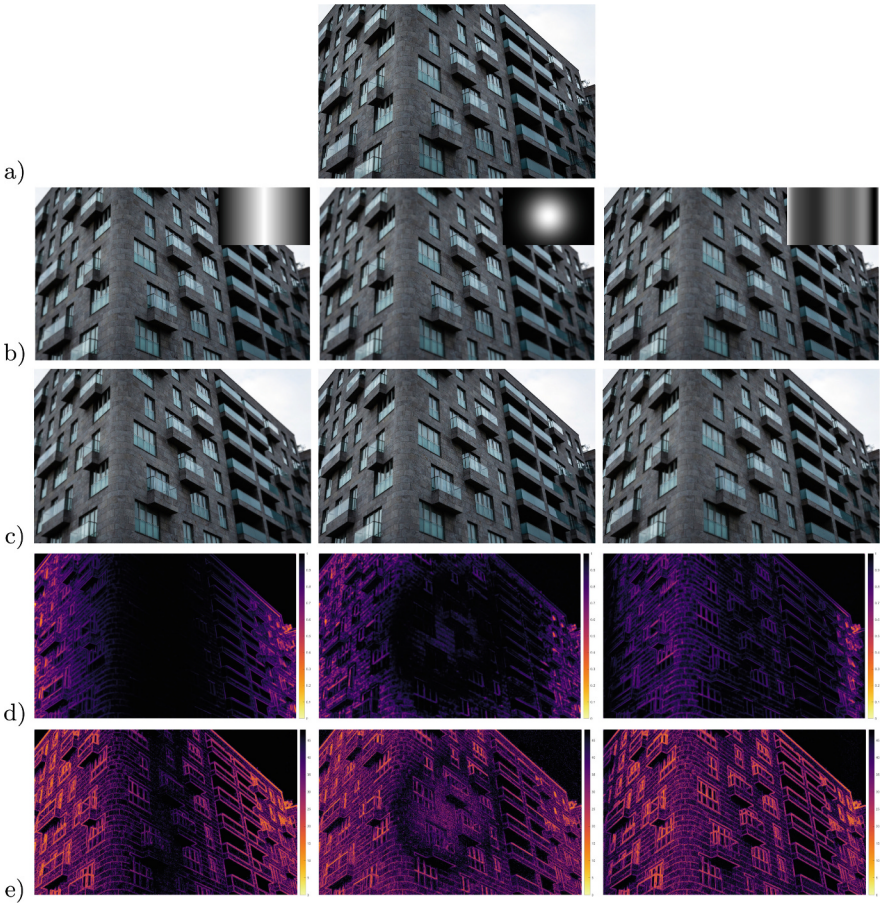


Fig. 2. Example result: a) original image; b) images degraded with their respective masks (linear, radial, and random); c) images restored with SPAN-s; d) SSIM value maps obtained by the original image and the restored one; e) the respective PSNR value maps. For all maps, a brighter color means larger errors, i.e. a lower level of details.

model to check the consistency of the results: Yolo8n, Yolo8s and Yolo8m. These models were fine tuned for 20 epochs using the same training configuration, with a learning rate of 0.01 and AdamW as the optimizer. Training was conducted on three different sets: the original images, the degraded images, and the images restored using the Super-Resolution methods. The top part of Table 3 reports the results obtained by the YOLO models on the restored test set,comparing their performance with models trained on the original (unmodified) images, degraded images, and restored images. The results are reported in terms of: Precision, indicating how many detections were correct; Recall, indicating the ratio of the instances identified by the model with respect to all instances of objects; mAP50,

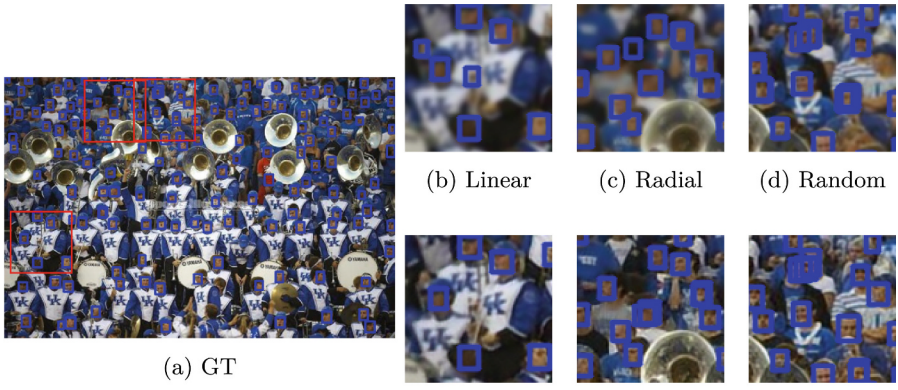
Table 3. Results obtained by the different versions of YOLO on the considered WIDER FACE Test set. Top table: the *Original* rows contain the values obtained by training the different versions of YOLO on the original images (without the degradation); the *Degraded* rows contain the results obtained by the YOLO models trained on the degraded images; the *SPAN* and *SPAN-s* rows contain the results obtained by the YOLO models trained on the restored images using the respective model. Bottom table: results obtained by testing the versions of YOLO trained using the original images on the degraded images and on the images restored using the SPAN and SPAN-S models. In bold are marked the best results obtained by the models trained on the restored or degraded images, while the second best are underlined.

	Model	Linear				Radial				Random			
		P	R	mAP50	mAP50-95	P	R	mAP50	mAP50-95	P	R	mAP50	mAP50-95
Original	Yolo8n	**0.783**	**0.480**	**0.561**	**0.282**	0.783	**0.480**	**0.561**	**0.282**	0.783	**0.480**	**0.561**	**0.282**
Degraded	Yolo8n	0.740	0.455	0.518	0.255	0.737	0.450	0.519	0.252	**0.785**	0.439	0.518	0.255
SPAN	Yolo8n	0.756	0.471	0.544	0.271	**0.788**	0.463	0.544	0.268	0.774	0.476	0.546	0.265
SPAN-s	Yolo8n	0.767	0.453	0.530	0.259	0.772	0.447	0.530	0.263	0.778	0.459	0.533	0.259
Original	Yolo8s	**0.795**	**0.508**	**0.603**	**0.316**	0.795	**0.508**	**0.603**	**0.316**	0.795	**0.508**	**0.603**	**0.316**
Degraded	Yolo8s	0.779	0.460	0.546	0.282	0.797	0.460	0.555	0.281	**0.881**	0.452	0.550	0.283
SPAN	Yolo8s	0.782	0.494	0.586	0.301	**0.805**	0.499	0.594	0.306	0.789	0.490	0.582	0.300
SPAN-s	Yolo8s	0.790	0.488	0.574	0.295	0.797	0.487	0.583	0.296	0.771	0.481	0.565	0.291
Original	Yolo8m	**0.825**	**0.514**	**0.618**	**0.331**	**0.825**	**0.514**	**0.618**	**0.331**	**0.825**	**0.514**	**0.618**	**0.331**
Degraded	Yolo8m	0.786	0.486	0.571	0.298	0.807	0.466	0.568	0.299	0.770	0.477	0.564	0.292
SPAN	Yolo8m	0.794	0.505	0.589	0.309	0.806	0.507	0.604	0.317	0.804	0.502	0.590	0.306
SPAN-s	Yolo8m	0.783	0.496	0.591	0.311	0.814	0.489	0.593	0.312	0.779	0.493	0.580	0.303

	Model	Linear				Radial				Random			
		P	R	mAP50	mAP50-95	P	R	mAP50	mAP50-95	P	R	mAP50	mAP50-95
Original	Yolo8n	0.783	**0.480**	**0.561**	**0.282**	**0.783**	**0.480**	**0.561**	**0.282**	**0.783**	**0.480**	**0.561**	**0.282**
Degraded	Yolo8n	0.757	0.447	0.525	0261	0.762	0.426	0.506	0.251	0.756	0.437	0.517	0.255
SPAN	Yolo8n	**0.785**	0.463	0.549	0.274	0.779	0.479	0.553	0.276	0.768	0.479	0.550	0.275
SPAN-s	Yolo8n	0.749	0.474	0.544	0.271	0.760	0.470	0.543	0.270	0.767	0.456	0.537	0.266
Original	Yolo8s	0.795	**0.508**	**0.603**	**0.316**	0.795	**0.508**	**0.603**	**0.316**	0.795	**0.508**	**0.603**	**0.316**
Degraded	Yolo8s	0.816	0.440	0.551	0.285	0.772	0.431	0.535	0.275	0.761	0.447	0.542	0.276
SPAN	Yolo8s	**0.822**	0.477	0.587	0.306	**0.819**	0.486	0.594	0.308	**0.802**	0.488	0.587	0.304
SPAN-s	Yolo8s	0.806	0.475	0.577	0.298	0.766	0.492	0.579	0.299	0.774	0.472	0.564	0.290
Original	Yolo8m	**0.825**	**0.514**	**0.618**	**0.331**	**0.825**	0.514	**0.618**	**0.331**	**0.825**	**0.514**	**0.618**	**0.331**
Degraded	Yolo8m	0.807	0.472	0.565	0.298	0.794	0.464	0.559	0.290	0.799	0.459	0.559	0.291
SPAN	Yolo8m	0.812	0.501	0.603	0.321	0.797	**0.517**	0.611	0.324	0.797	0.504	0.601	0.316
SPAN-s	Yolo8m	0.810	0.488	0.593	0.313	0.793	0.502	0.596	0.313	0.821	0.465	0.577	0.305

i.e. mean average precision calculated at an intersection over union (IoU) threshold of 0.50; mAP50-95, i.e. the average of the mean average precision calculated at varying IoU thresholds, ranging from 0.50 to 0.95 [22]. Since we are interested in the overall assessment of model performance, the reference metric to consider is mAP. The obtained results show that, in all observed cases, models trained on images enhanced by a Super-Resolution method outperform those trained directly on degraded images.

More specifically, models incorporating the trained Super-Resolution methods to correct degradations achieve higher *mAP* values than those that do not, though their performance remains slightly below that of models trained on the original images. Figure 3 provides visual examples of face detections using SPAN-s for restoration and Yolo8n for detection. These examples highlight that applying restoration methods leads to more accurate detections and an increased number of detected faces. The bottom part of Table 3 also reports



(a) GT (b) Linear (c) Radial (d) Random

(e) Linear SPAN-s (f) Radial SPAN-s (g) Random SPAN-s

Fig. 3. Results obtained by using Yolo8n in the various analyzed cases. Where there is a stronger degradation (e.g., the areas in the red boxes) the methods that use the restored images detect more faces or are more accurate in localization. (Color figure online)

the results of YOLO models trained on original images and tested on both degraded and restored versions. In these cases, the models consistently achieve better performance on enhanced images than on degraded ones. This indicates that, even when trained solely on unmodified images, the models benefit from Super-Resolution preprocessing, improving detection performance despite not fully recovering all lost details.

5 Conclusions

This paper presents an investigation of the use of Efficient Super-Resolution (ESR) techniques to restore images with spatially varying degradations. By leveraging lightweight ESR models, we demonstrated their effectiveness in recovering image details and enhancing the visual quality of degraded images, particularly in scenarios with uneven resolution artifacts. Furthermore, we showcased the practical impact of ESR-restored images on downstream tasks, such as object detection, using the WIDER FACE dataset in conjunction with YOLOv8. The results underline how applying ESR methods significantly improves object detection performance compared to models operating on degraded images, though there is still a gap compared to models trained on pristine data.

While ESR models offer a computationally efficient solution, their performance is currently surpassed by more advanced methods, such as Transformer-based super-resolution models (e.g., SwinIR). Future work could focus on integrating lightweight Transformer-based architectures to better handle spatially varying degradations while maintaining efficiency. Another key limitation of this study is the reliance on synthetic datasets to simulate varying resolutions due

to the lack of real-world datasets containing such degradations. Developing a dedicated real-world dataset, including images captured with devices prone to spatially varying degradations, such as 360° cameras, fish-eye lenses, or sonars, would provide a more robust basis for evaluation. As future work we plan to extend our method to other application domains, such as medical imaging, remote sensing or autonomous driving, which often encounter spatially varying resolution issues due to the nature of sensors used (e.g., sonars, radars, lidars, optical and electron microscopes), and where enhanced resolution may significantly improve interpretability and decision-making. We also plan to extend our method without making any assumptions about the degradation kernel thus permitting to experiment with real data even in absence of a higher resolution groundtruth.

Acknowledgments. This work has been carried out within the High-Performance Computing, Big Data and Quantum Computing National Research Center project, funded by the European Union - NextGenerationEU, under the National Recovery and Resilience Plan (NRRP) Mission 4 Component 2 Investment Line 1.4, project CN_00000013 / CN1, Innovation Grant SURE, CUP H43C22000520001.

References

1. Agustsson, E., Timofte, R.: Ntire 2017 challenge on single image super-resolution: dataset and study. In: Proceedings of the IEEE Conference on CVPR Workshops, pp. 126–135 (2017)
2. Aljadaany, R., Pal, D.K., Savvides, M.: Douglas-Rachford networks: learning both the image prior and data fidelity terms for blind image deconvolution. In: Proceedings of the IEEE/CVF Conference on CVPR, pp. 10235–10244 (2019)
3. Bevilacqua, M., Roumy, A., Guillemot, C., Alberi-Morel, M.L.: Low-complexity single-image super-resolution based on nonnegative neighbor embedding (2012)
4. Dong, C., Loy, C.C., He, K., Tang, X.: Learning a deep convolutional network for image super-resolution. In: Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T. (eds.) ECCV 2014. LNCS, vol. 8692, pp. 184–199. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-10593-2_13
5. Esmaeilzehi, A., Ahmad, M.O., Swamy, M.: UPDCNN: a new scheme for image upsampling and deblurring using a deep convolutional neural network. In: 2019 IEEE International Conference on Image Processing (ICIP), pp. 2154–2158 (2019)
6. Esmaeilzehi, A., Ahmad, M.O., Swamy, M.: UPDRESNN: a deep light-weight image upsampling and deblurring residual neural network. IEEE Trans. Broadcast. **67**(2), 538–548 (2021)
7. Goodfellow, I., et al.: Generative adversarial networks. Commun. ACM **63**(11), 139–144 (2020)
8. Huang, J.B., Singh, A., Ahuja, N.: Single image super-resolution from transformed self-exemplars. In: Proceedings of the IEEE Conference on CVPR, pp. 5197–5206 (2015)
9. Jocher, G., Chaurasia, A., Qiu, J.: Ultralytics YOLOv8 (2023). https://github.com/ultralytics/ultralytics
10. Kong, F., Li, M., et al.: Residual local feature network for efficient super-resolution. In: Proceedings of the IEEE/CVF Conference on CVPR, pp. 766–776 (2022)

11. Kupyn, O., Budzan, V., Mykhailych, M., Mishkin, D., Matas, J.: DeblurGAN: blind motion deblurring using conditional adversarial networks. In: Proceedings of the IEEE Conference on CVPR, pp. 8183–8192 (2018)
12. Laroche, C., Almansa, A., Tassano, M.: Deep model-based super-resolution with non-uniform blur. In: Proceedings of the IEEE/CVF WACV, pp. 1797–1808 (2023)
13. Ledig, C., Theis, L., et al.: Photo-realistic single image super-resolution using a generative adversarial network. In: Proceedings of the IEEE Conference on CVPR, pp. 4681–4690 (2017)
14. Liang, J., Cao, J., et al.: Swinir: Image restoration using swin transformer. In: Proceedings of the IEEE/CVF ICCV, pp. 1833–1844 (2021)
15. Liu, J., Tang, J., Wu, G.: Residual feature distillation network for lightweight image super-resolution. In: Bartoli, A., Fusiello, A. (eds.) ECCV 2020. LNCS, vol. 12537, pp. 41–55. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-67070-2_2
16. Liu, Z., et al.: Swin transformer: hierarchical vision transformer using shifted windows. In: Proceedings of the IEEE/CVF ICCV, pp. 10012–10022 (2021)
17. Martin, D., Fowlkes, C., Tal, D., Malik, J.: A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In: Proceedings of the IEEE ICCV, vol. 2, pp. 416–423. IEEE (2001)
18. Matsui, Y., Ito, K., et al.: Sketch-based manga retrieval using manga109 dataset. Multimedia Tools Appl. **76**, 21811–21838 (2017)
19. Mustaniemi, J., Kannala, J., Särkkä, S., Matas, J., Heikkila, J.: Gyroscope-aided motion deblurring with deep networks. In: 2019 IEEE Winter Conference on Applications of Computer Vision (WACV), pp. 1914–1922. IEEE (2019)
20. Nagy, J.G., O'Leary, D.P.: Restoring images degraded by spatially variant blur. SIAM J. Sci. Comput. **19**(4), 1063–1082 (1998)
21. Nah, S., Hyun Kim, T., Mu Lee, K.: Deep multi-scale convolutional neural network for dynamic scene deblurring. In: Proceedings of the IEEE Conference on CVPR, pp. 3883–3891 (2017)
22. Padilla, R., Netto, S.L., Da Silva, E.A.: A survey on performance metrics for object-detection algorithms. In: 2020 International Conference on Systems, Signals and Image Processing (IWSSIP), pp. 237–242. IEEE (2020)
23. Ren, B., Li, Y., et al.: The ninth ntire 2024 efficient super-resolution challenge report. In: Proceedings of IEEE/CVF Conference on CVPR, pp. 6595–6631 (2024)
24. Ren, D., Zhang, K., Wang, Q., Hu, Q., Zuo, W.: Neural blind deconvolution using deep priors. In: Proceedings of IEEE/CVF Conference on CVPR, pp. 3341–3350 (2020)
25. Schuler, C.J., Hirsch, M., Harmeling, S., Schölkopf, B.: Learning to deblur. IEEE Trans. Pattern Anal. Mach. Intell. **38**(7), 1439–1451 (2015)
26. Shi, W., Caballero, J., et al.: Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. In: Proceedings of the IEEE Conference on CVPR, pp. 1874–1883 (2016)
27. Wan, C., Yu, H., et al.: Swift parameter-free attention network for efficient super-resolution. In: Proceedings of the IEEE/CVF Conference on CVPR, pp. 6246–6256 (2024)
28. Yang, S., Luo, P., Loy, C.C., Tang, X.: Wider face: a face detection benchmark. In: Proceedings of the IEEE Conference on CVPR, pp. 5525–5533 (2016)
29. Yang, W., Zhang, X., et al.: Deep learning for single image super-resolution: a brief review. IEEE Trans. Multimedia **21**(12), 3106–3121 (2019)

30. Yu, L., Li, X., et al.: Dipnet: efficiency distillation and iterative pruning for image super-resolution. In: Proceedings of the IEEE/CVF Conference on CVPR, pp. 1692–1701 (2023)
31. Yue, L., Shen, H., Li, J., Yuan, Q., Zhang, H., Zhang, L.: Image super-resolution: the techniques, applications, and future. Sig. Process. **128**, 389–408 (2016)
32. Zeyde, R., Elad, M., Protter, M.: On single image scale-up using sparse-representations. In: Curves and Surfaces: 7th International Conference, pp. 711–730 (2012)
33. Zhang, K., et al.: Deep image deblurring: a survey. IJCV **130**(9), 2103–2130 (2022)