

Comparison of Contemporary Image Resolution Upscaling Algorithms: a Survey

Anuar Nurlybayev

Robotics and Mechatronics department Nazarbayev University Nur-Sultan, Kazakhstan anuar.nurlybayev@nu.edu.kz

Andrey Yershov

Robotics and Mechatronics department Nazarbayev University Nur-Sultan, Kazakhstan andrey.yershov@nu.edu.kz

Maxat Tassemenev

Robotics and Mechatronics department Nazarbayev University Nur-Sultan, Kazakhstan maxat.tassemenev@nu.edu.kz

Index Terms—image super-resolution, image processing, upscaling, deep convolutional network, sparse representation, dictionaries, image restoration, adaptation model, image reconstruction, deblurring, computational efficiency

Abstract—Image Super-Resolution (SR) is a class of image processing techniques to enhance the resolution of images and videos in computer vision tasks. Modern image super-resolution techniques have different implementation methods and accordingly the results' quality, performance time and consumed computational power have distinctions. This survey paper examines three main super-resolution approaches based on Deep Learning, Dictionaries and Self-Similarity. Comparison of output images' quality of tested algorithms was based our own dataset of 5 images of different content and resolution. The survey concluded that deep network based super-resolution algorithm provides the best results among contemporary upscaling methods.

I. INTRODUCTION

Digital images have become a indispensable standard for people's daily visual communication and a method to preserve data. However, digital images have limited resolution, which is often not enough for comfortable usage and acquisition of necessary details of the digital image. Image resolution upscaling, or single image super-resolution (SR) technique is used to obtain a higher-resolution image or sequence from the observed lower-resolution images. This involves obtainment of higher number of pixels, which have to be represented on a larger bitmap and sampled from original smaller grid of pixels. Super-resolution is a technique to achieve the best image from degraded low-resolution image or images. Super-resolution consists of image interpolation and image restoration. Possible restoration may be needed to restore the changed dimension of the image due to the image interpolation [1] [2].

The main issue of resolution upscaling is that simple, non-adaptive techniques such as nearest-neighbor interpolation, bilinear and bicubic interpolations usually do not deliver satisfactory, natural looking result. The upscaled image using these algorithms is usually distorted, blurry and/or lacks certain details or has clearly visible unnatural, rough edges, all of which are unwanted artifacts of simple algorithms. Recently, a number of adaptive techniques (the ones that act differently for every image, according to content) have been developed. The objective of this paper is to provide an overview of different method approaches and compare their results.

This topic has recently become relevant due to emergence of available high resolution screens in our daily electronics: smartphones, TVs, laptops, PCs. Satellite imaging needs reliable image resolution upscaling in order to track ground events for fields such as navigation, agriculture, forest fire protection and mitigation. Graphic designers and photographers need image resolution upscale for obtaining better prints and higher amount of details and sharpness in their work. Contemporary autonomous driving systems and robots usually use low resolution vision systems in order to save amount of transmitted and saved data, however, the acquired data could require resolution upscale, which is both not resource intensive to accomplish and produces result that would not take up a lot of memory. All in all, every modern field which involves image and video products will always be in need for a reliable algorithm of image resolution upscaling which produces high quality result for every given task.

II. IMAGE QUALITY ASSESSMENT

This section is aimed on discussion of methods of resultant image quality assessment (IQA). There are a number of parameters which have to be assessed while evaluating and comparing image resolution upscaling algorithms. The most important one is certainly the quality of the resultant image. The processing time, amount of memory allocated for the process and algorithm's dependence on hardware are also considered. Optimal image resolution upscaling algorithm has to balance time and quality. Appearance of undesirable results such as blur, compression artifacts, false colours are going to be assessed on different methods as well. While the aim is to produce a result, which looks detailed yet natural, there should be a mathematical assessment tool to distinguish not obvious differences in resultant images. The reason for this is that human perception of digital image could be biased and is prone to optical illusions. It depends upon various factors, such as background color of the observed image, quality and characteristics of the screen, human's eye health etc. For this purpose, a number of IQA tests were conducted on the resultant images from SR algorithms. The IQA results with instructions how to read them could be found in Appendix H. There are full reference based methods and no-reference methods available. While reference based methods provide more

accurate assessment through comparison of original reference and the input pixel by pixel, its main disadvantage is that it could be easily deceived when there are distortions involved in the investigated image or it has mismatching pixel count with the original image. This, however, is not an issue with the no-reference methods. Their drawback is that classical, statistics based no-reference IQA methods do not represent human perception of the image to the extent that deep learning based no-reference IQA methods do. ML based IQA were not implemented in this study, however, they are considered to be the feature standard for image quality assessment. Regarding the IQA methods, the first one of no-reference algorithms is peak signal-to-noise ratio (PSNR), which is derived from the mean square error, and indicates the ratio of the maximum pixel intensity to the power of the distortion [3].

The next two methods - Structural similarity index (SSIM) and Feature similarity index (FSIM) have similar nature - it combines local image structure, luminance, and contrast into a single local quality score. While SSIM is based on structures, FSIM uses features, and is believed to be enhanced version of SSIM [4]. The no-reference Natural Image Quality Evaluator (NIQE) model is trained on a database of undistorted images, and the Perception based Image Quality Evaluator (PIQE) is opinion-unaware and unsupervised, which means it does not require a trained model. PIQE estimates block-wise distortion and measures the local variance of perceptibly distorted blocks to compute the quality score [3].

III. METHODOLOGY

Three approaches of SR were considered during this paper: deep learning-based super-resolution, dictionary-based super-resolution and self-similarity-based super-resolution.

A. Deep Learning-based Super-Resolution

Deep learning based SR is a method that uses neural networks to map low-resolution segments to high resolution output. This models have been actively researched and often achieve the state-of-the-art performance. Several learning methods have been applied to solve SR, starting from Convolutional Neural Networks (CNN) to the most recent Generative Adversarial Nets. In general, deep learning-based SR algorithms differ from each other by the types of network architectures, learning principles and loss functions [5]. Enhanced Deep Residual Network (EDSR) is an example of deep learning SR methods that uses residual multi-scale model that is able to reconstructs high-resolution images for various scales. Lim et.al. optimized the model by removing unnecessary modules in conventional residual networks and expanded the model size during the training procedure [6]. EDSR uses post-upsampling SR network that performs most computation in low-dimensional space by replacing the predefined upsampling with end-to-end learnable layers integrated at the end of the models. EDSR has became one of the basis of CNN SRs, and it is considered as a state-of-the-art SR model [5]. The network architecture can be seen in the Figure 1. Lai et.al. proposed an improvement to the Laplacian pyramid method and used Deep

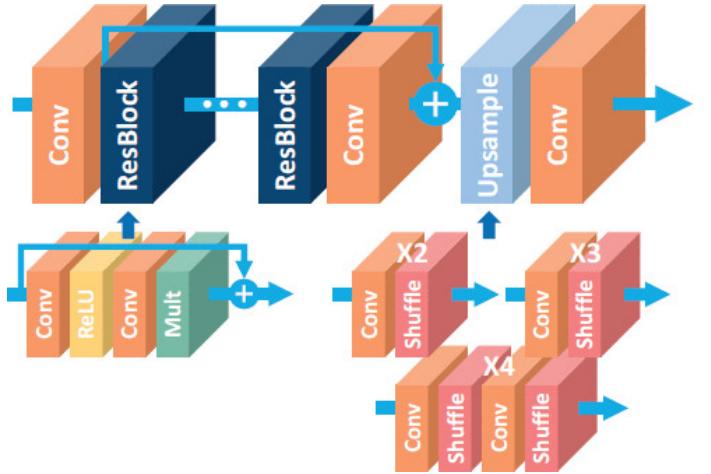


Fig. 1. Network architecture of the Enhanced Deep Residual Networks [6]

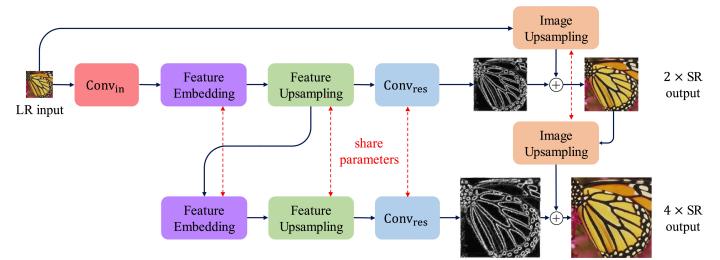


Fig. 2. Network architecture of the Deep Laplacian Pyramid Network [7]

Laplacian pyramid network for super-resolution. The network uses progressive upsample approach, where features of the low resolution image are extracted through convolution. These features collapse into the upsampled image and go through next level of the network. This method allowed to improve multi-scale training that can handle multiple upscaling scopes with a single model, parameter sharing, processing speed and progressive reconstruction. In addition, this method avoids pre-processing bicubic interpolation and directly extracts features from the low-resolution input [7]. The network architecture can be seen in the Figure 2.

B. Dictionary-based Super-Resolution

Dictionary-based SR is super-resolution approach that relies on dictionaries of image patches or patch-based atoms.

Neighbors embedding SR is a more generalized approach that approximates all patches that lie on manifolds in their respective LR and HR spaces, which allows for an input patch to be approximated as an interpolation of existing database patches, with one set of interpolation coefficients being applied to both spaces. Anchored Neighborhood Regression (ANR) is one of the fastest SR methods that achieves state-of-the-art results [8]. Timofte et.al. proposed new Adjusted Anchored Neighborhood Regression (A+) that combines simple functions of sparsity-based SR along with ANR, to improve the time complexity of the sparsity-based SR [8]. We chose this

paper because it proposed an advancement in the dictionary-based SR.

Sparsity based SR relies on sparse representation of image patches. By using sparsity constraints, it is possible to represent low resolution patches and their high resolution analogues through a sparse representation. Trained sparse representation from the dictionaries are used to improve low-resolution input images through comparison. This representation has an invariance assumption that the patch pairs have the same space representation [9]. Peleg et.al. proposed a model (SISR) that utilizes the statistical dependencies between the sparsity patterns of the LHR coefficients and between the corresponding nonzero coefficients, which allows the model to use dictionaries of different sizes. To improve the performance, data clustering was added along with multi-level scale-up scheme [10]. Adaptive sparse domain selection with adaptive regularization (ASDS_AR) can be applied to the sparse representation to improve image restoration, which was implemented by Dong et.al [11].

C. Self-similarity-based Super-Resolution

Most single image SR methods rely on external databases of LR-HR patches pairs. Huang et.al. proposed self-similarity driven SR algorithm (SelfExSR) that is able to build patch pairs without databases. Huang et.al expanded the internal patch search space by allowing geometric variations through explicit localizing planes inside the scene, the patch search process was improved by using the detected geometry to guide the search. The compositional model allowed to handle generic and additional affine transformations to accommodate local shape variation [12].

IV. RESULTS

The performance of different methods can be compared within three criteria: the performance time, the computational power and the quality of the image. Deep learning based SR have small performance time, with an average of 50.8 s. (average EDSR - 30.8s, LapSRN - 70.8s). The performance time of the ANR was 13.5 s., which was expected as it is one of the fastest methods. Sparsity-based approaches require some time to perform with an average of 15 minutes, however Peleg et.al. method improved the time to 5 minutes. ASDS_AR peak performance was at 13 minutes, which is also an improvement relative to sparse general functions SR. A+ method has a computation time of 86 s., which is slower than ANR, but is much faster than sparsity-based SR. The computation time of the SelfEx was the highest at 15+ minutes, because the algorithm performs super-resolution without any external data. The performance time may differ as it is very dependent on the work station conditions.

The computational power required to perform varies for different approaches. Deep Learning SR require good performance, as they were the only one who performed using a graphic card with CUDA. Convolution networks were implemented and tested in Ubuntu 16.04 using PyTorch. Several different deep learning-based SR were considered, however

with a drop of support of the MATLAB based convolution networks they are no longer reproducible. The training process of the model varies between 1-2 hours, depending on the working station performance, while other interpolation techniques were performed in MATLAB and were dependent only on the MATLAB performance. Dictionary-based SR's main disadvantage is that it relies on a pre-trained dictionary of a low- and high-patches to work. The dictionary requires a large database of image patches and an increased amount of atoms with respect to patches, dictionary training has a high time complexity. Tested dictionaries of the methods were identical as the dictionaries used in the corresponding papers. Another disadvantage is that external image restoration may be needed as the output image dimensions will not exactly match with the original HR image. This issue is encountered because of numerous divisions of spatial resolution values, which results in lost pixels in case the value is odd. Ultimately, this issue affects reference based image quality assessment score. The performance of the approaches can be seen in the Figures 3 and 4.

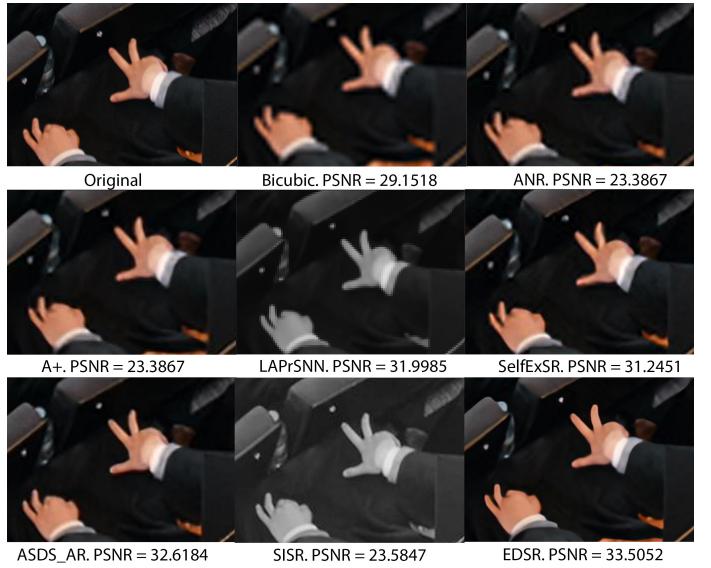


Fig. 3. Comparison of results on "Graduates"

As it could be seen from results above, EDSR has shown the most impressive result that resembles original image. The resultant images are sharp and clear, without any artifacts or noise present. Its PSNR score on "graduates" figure confirms this as well. However, PSNR score of the same method on the figure 4, "Expo", has shown much lower score. This is also true for other SR methods for this figure as well. This could be due to nature of the figure - there are considerable amount of tiny detail, which are lost during down-sampling. Nevertheless, visually EDSR has shown the best performance on this figure as well. The second best performance was achieved by ASDS method. Its sharpness and clearness is a little lower than that of the first method, with some of the color noise present at the edges of objects. The third best method by the PSNR score is LAPrSNN method. This method has shown the best

performance on IQA test on "Expo" figure. However, visual assessment of this method's output does not coincide with these scores on mathematical quality assessment. As it could be seen from the figures above and in appendix C, LAPrSNN results lack much of the details and are have much more pronounced pixel edges on objects, as if the nearest-neighbor interpolation was performed.

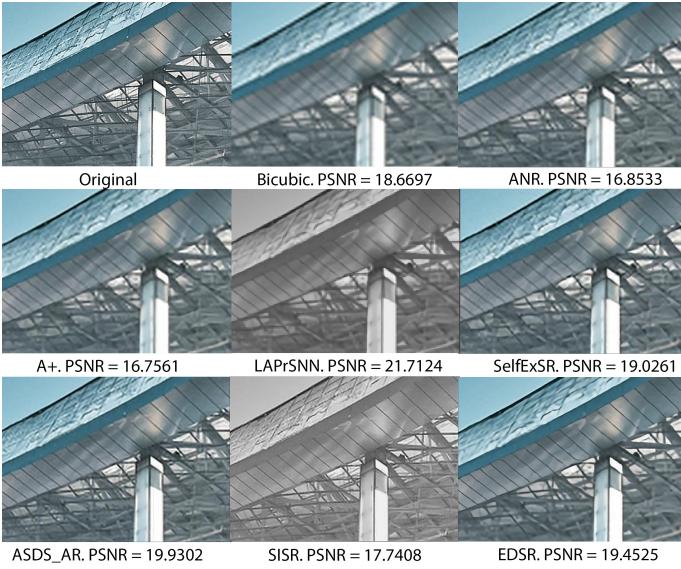


Fig. 4. Comparison of results on "Expo"

Based on our visual assessment, SISR method has shown very sharp and clear output, with much of the original detail recovered without noise. However, it failed nearly all of the IQA tests. This could be due to pixel shifts during SR algorithm application, ultimately causing misalignment between the original image and the output during comparison by reference based IQA tests. While this explanation is true for reference based tests, non-reference based IQA tests for the SISR outputs has produced low scores as well. This clearly shows flaws in the IQA tests, as long as the output quality and the IQA scores does not have a correlation at this case. As of Bicubic interpolation, it was included in the results for comparison purposes, and it clearly loses to all of the methods tested, with fuzzy edges of the objects and lost details all over the image. However, its IQA results have much higher scores than some of the methods at some of the images, what again does not co-relate with visual assessment of the results.

Regarding other methods' results - A+, ANR and SelfExSR, the resultant image performance was nearly on the same level for all of them. Considerable loss of detail was noticed, with edges lacking sharpness overall. SelfExSR output images have shown some artifacts - small parts of the output images had modifications that were not present at the original image. For instance, in the "Graduates" figure SelfExSR algorithm has changed the shape of one of the human's fingers in the lower

left part of the image. Similar problem could be noticed on the figure "Men" in the middle left part in appendix B.

V. CONCLUSION

This paper's focus is on the survey of the single image super-resolution methods. Single image SR could be executed using deep learning-based methods, dictionary-based methods and self-similarity based methods. Outputs of all of the methods were tested on a dataset of 5 images down scaled bicubically. Overall, all approaches reached state-of-the-art performance, with the best results provided by deep learning SR with post-upsampling residual network and sparsity-based SR with adaptive sparse domain selection.

REFERENCES

- [1] "Image Super Resolution-A Survey, author=Amisha J. Shah and Suryakant B. Gupta," in *2012 1st International Conference on Emerging Technology Trends in Electronics, Communication and Networking*, 2012.
- [2] C. S. Balure and M. R. Kini, "A Survey - Super Resolution Techniques for Multiple, Single, and Stereo Images," in *2014 Fifth International Symposium on Electronic System Design*, 2014.
- [3] "MATLAB. Image Quality Metrics - MATLAB Simulink," <https://www.mathworks.com/help/images/image-quality-metrics.html>., accessed: April 25, 2020.
- [4] L. Zhang, L. Zhang, X. Mou, and D. Zhang, "FSIM: A Feature Similarity Index for Image Quality Assessment," *IEEE Transactions on Image Processing*, vol. 20, no. 8, p. 2378–2386, 2011.
- [5] Z. Wang, J. Chen, and S. C. Hoi, "Deep Learning for Image Super-resolution: A Survey," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2020.
- [6] B. Lim, S. Son, H. Kim, S. Nah, and K. M. Lee, "Enhanced Deep Residual Networks for Single Image Super-Resolution," in *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2017.
- [7] W.-S. Lai, J.-B. Huang, N. Ahuja, and M.-H. Yang, "Fast and Accurate Image Super-Resolution with Deep Laplacian Pyramid Networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 41, no. 11, pp. 2599 – 2613, 2019.
- [8] R. Timofte, V. D. Smet, and L. V. Gool, "A+: Adjusted Anchored Neighborhood Regression for Fast Super-Resolution," in *Computer Vision – ACCV 2014 : 12th Asian Conference on Computer Vision, Singapore, Singapore, November 1-5, 2014, Revised Selected Papers, Part IV*, A. Koubaa, Ed. Springer, 2015, pp. 111–126.
- [9] J. Yang, J. Wright, T. S. Huang, and Y. Ma, "Image Super-Resolution Via Sparse Representation," *IEEE Transactions on Image Processing*, vol. 19, no. 11, pp. 2861 – 2873, 2010.
- [10] T. Peleg and M. Elad, "A Statistical Prediction Model Based on Sparse Representations for Single Image Super-Resolution," *IEEE Transactions on Image Processing*, vol. 23, no. 6, pp. 2569 – 2582, 2014.
- [11] W. Dong, L. Zhang, G. Shi, and X. Wu, "Image Deblurring and Super-resolution by Adaptive Sparse Domain Selection and Adaptive Regularization," *IEEE Transactions on Image Processing*, vol. 20, no. 7, pp. 1838 – 1857, 2011.
- [12] J.-B. Huang, A. Singh, and N. Ahuja, "Single image super-resolution from transformed self-exemplars," in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015.

APPENDIX A
EDSR RESULTS

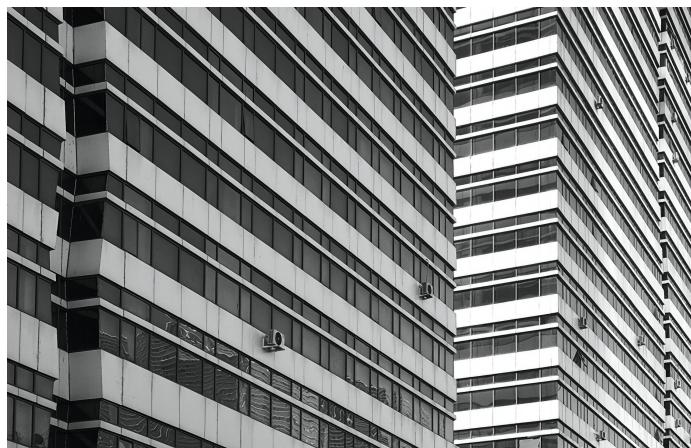


Fig. 5. Result on "Wall"

APPENDIX B
SELFEXSR RESULTS

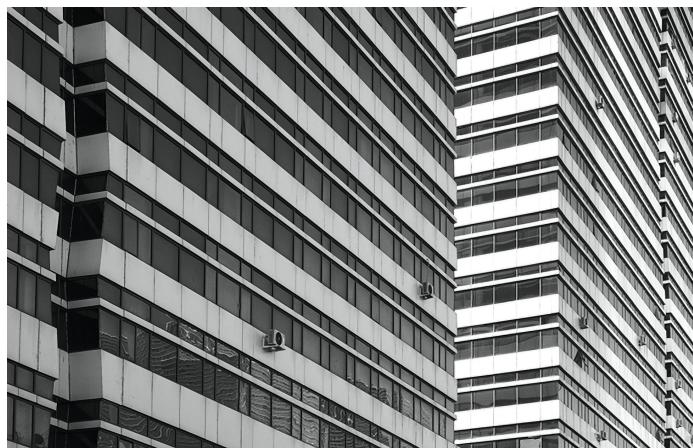


Fig. 8. Result on "Wall"



Fig. 6. Result on "Sun"



Fig. 9. Result on "Sun"



Fig. 7. Result on "Men"

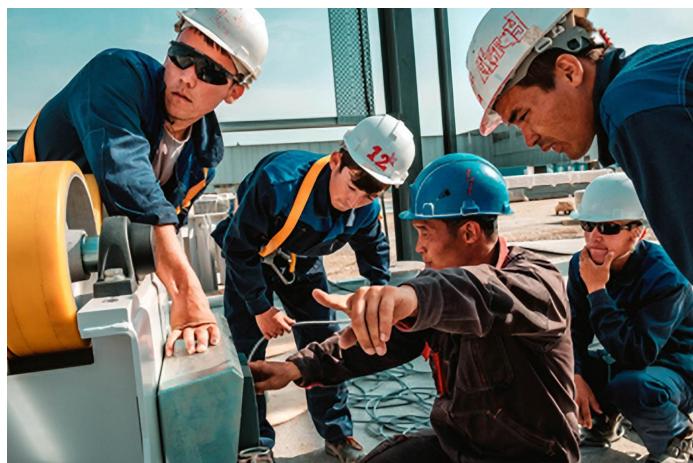


Fig. 10. Result on "Men"

APPENDIX C
LAPrSNN RESULTS

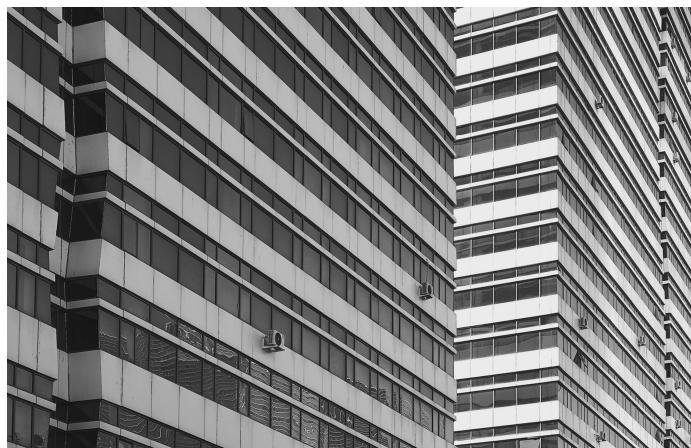


Fig. 11. Result on "Wall"

APPENDIX D
SiSR RESULTS



Fig. 14. Result on "Wall"

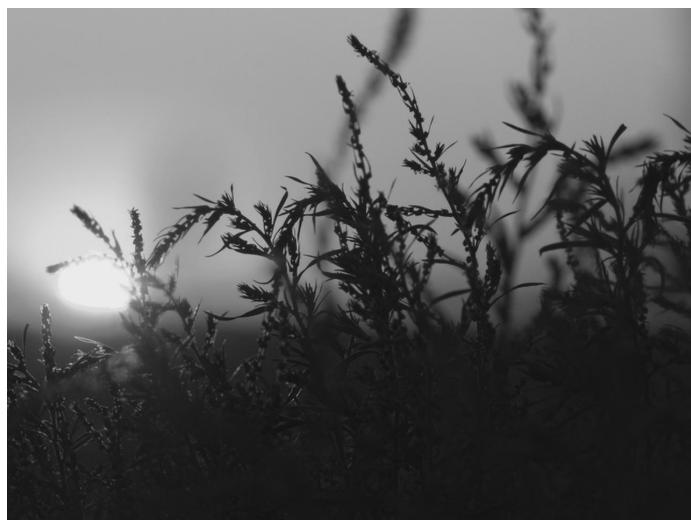


Fig. 12. Result on "Sun"

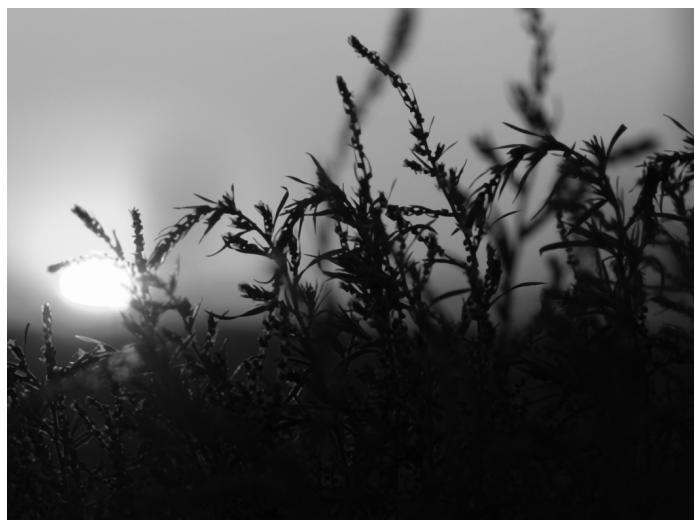


Fig. 15. Result on "Sun"



Fig. 13. Result on "Men"



Fig. 16. Result on "Men"

APPENDIX E
A+ RESULTS

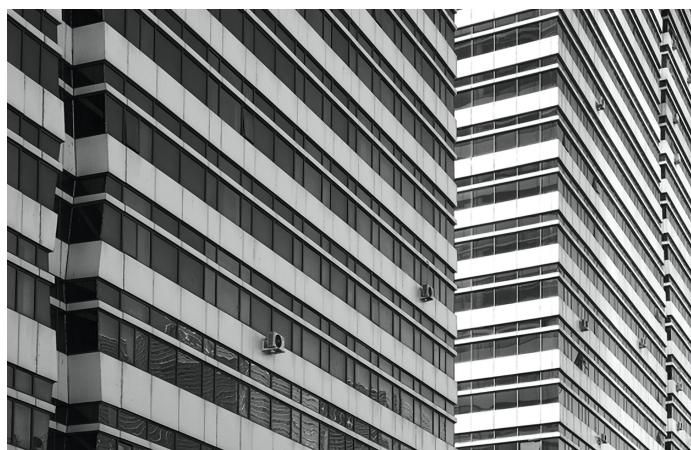


Fig. 17. Result on "Wall"

APPENDIX F
ANR RESULTS

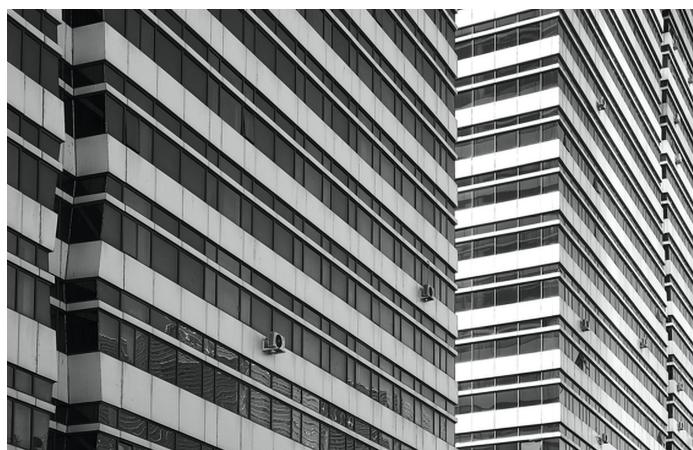


Fig. 20. Result on "Wall"



Fig. 18. Result on "Sun"



Fig. 21. Result on "Sun"



Fig. 19. Result on "Men"



Fig. 22. Result on "Men"

APPENDIX G ASDS RESULTS



Fig. 23. Result on "Wall"



Fig. 24. Result on "Sun"



Fig. 25. Result on "Men"

APPENDIX H IQA PERFORMANCES

The leftmost column shows the name of the method and names of the images tested. All of the following columns feature the name of Image Quality Assessment tests and results for the corresponding images. Red highlight resembles poor (lower than average) performance on the test, while green means that the test score is higher than average in the corresponding method of image quality assessment. White highlight is average performance.

	PSNR	SSIM	FSIM_grey	FSIM_RGB	pique	nique
EDSR						
expo	19.4525235	0.755240714	0.989856945	0.989794274	70.3073853	3.72662154
grads	33.5052134	0.951822338	0.998906319	0.998845097	72.9097521	4.73608937
man	27.4736044	0.953265965	0.992126217	0.991947457	74.4717212	3.80187635
sun	34.5290837	0.980656673	0.998182699	0.998080124	76.20994	4.38368846
wall	23.2028027	0.796272087	0.993205661	0.99317935	60.9410702	4.29071963
SelfExSR_IQA						
expo	19.0261614	0.727138992	0.988729733	0.988561484	47.9569638	3.43733285
grads	31.24509	0.925278786	0.998423154	0.998222792	50.4032691	4.16223427
man	26.1505314	0.905307177	0.990849704	0.99039672	60.3057053	3.28681429
sun	32.099239	0.968568029	0.997108633	0.996810225	56.6441711	3.79173338
wall	21.9110902	0.734331253	0.991502536	0.991466644	51.7824816	4.08697536
laprsnn_IQA						
expo	21.71246807	0.724041231	0.956950613	0.956950613	42.61781915	2.784837228
grads	31.99851512	0.923808769	0.99416702	0.99416702	63.66912586	4.55280745
man	25.37306161	0.884091511	0.982037213	0.982037213	77.00471101	3.976558523
sun	35.46967173	0.961973838	0.998054537	0.998054537	81.30996028	4.19528635
wall	19.66814851	0.764992341	0.940912258	0.940912258	68.83178934	4.637050858
SISR_IQA						
expo	17.74088246	0.544458164	0.948623715	0.948623715	24.76019775	3.11175194
grads	23.58475548	0.701038888	0.966803531	0.966803531	49.26067794	4.335734211
man	21.86993895	0.765519921	0.941990951	0.941990951	47.25283356	2.557643209
sun	25.70993355	0.831432671	0.97563418	0.97563418	51.09192214	3.737139366
wall	16.95760635	0.637144278	0.933991226	0.933991226	45.46716876	4.013798026
A+ AANR_IQA						
expo	16.75611169	0.651942089	0.890560614	0.890019427	53.05641406	3.368500639
grads	23.24227422	0.821305435	0.930131012	0.929075111	56.37221392	4.567933702
man	20.00230381	0.795142772	0.889780588	0.888000862	62.29951176	3.367407902
sun	26.61347422	0.939718246	0.955545326	0.954519377	59.25800692	3.80440994
wall	14.17301759	0.543265154	0.834960183	0.834927444	57.74600822	4.087281909
ASDS_AR_NL_TD1_IQA						
expo	19.9302885	0.776490795	0.993447717	0.993205054	38.2870298	3.2535762
grads	32.6184126	0.939441725	0.999266823	0.998794984	55.3555991	4.48666215
man	26.4904565	0.915609833	0.986719767	0.985911584	58.2524609	3.10163435
sun	31.4394291	0.969197053	0.99366136	0.993124888	56.4379969	3.89556673
wall	18.6992954	0.685752055	0.958847169	0.958811649	53.0710313	3.92276145
bicubic						
expo	18.6697005	0.695229683	0.985310969	0.985242477	73.0277649	4.25543346
grads	29.1518094	0.906677905	0.997104875	0.997010979	69.9172305	5.40572941
man	24.5450699	0.875200316	0.977885619	0.977583951	77.4476346	4.60489331
sun	31.2486666	0.972286113	0.994616559	0.994426149	79.7846758	4.83480925
wall	20.0319348	0.637214037	0.98531267	0.985288115	73.4406935	5.21787913
ANR						
expo	16.853327	0.651208348	0.890337655	0.889838246	58.4551151	3.6302298
grads	23.3867361	0.818400702	0.929948929	0.928892274	60.5190273	4.58246654
man	20.1324126	0.792873055	0.888724361	0.88694514	68.7796227	3.87531128
sun	26.6654419	0.939737401	0.955459236	0.954435154	61.6529961	3.96405126
wall	14.288031	0.538046815	0.834876094	0.834843416	63.7367559	4.56443793