FACULTY OF AUTOMATIC CONTROL, ELECTRONICS AND COMPUTER SCIENCE

DOCTORAL DISSERTATION SUMMARY

Multi-image super-resolution reconstruction using deep graph neural networks

---

mgr inż. Tomasz Tarasiewicz

Supervisor: **dr hab. inż. Michał Kawulok, prof. PŚ**

September 28, 2023

Chapter 1: Introduction

The significance of image resolution across various domains [9], such as medicine [5], satellite imaging [10], and digital photography [3], is underscored at the outset. However, the attainment of high-resolution (HR) images is hindered by certain physical and digital limitations [8, 11, 19]. Super-resolution reconstruction (SRR) emerges as a viable solution to augment the spatial resolution of existing low-resolution (LR) images.

The chapter delineates between single-image super-resolution (SISR) and multi-image super-resolution (MISR). SISR aims to enhance the resolution of a single image [7], while MISR endeavours to utilize multiple images to create a more detailed HR image [16]. The two primary stages of MISR, registration and fusion, are elaborated, along with the challenges like temporal differences, occlusions, and varying lighting conditions [12].

The limitations of traditional convolutional neural networks (CNNs) [14] for MISR are discussed, highlighting issues like handling temporal variations among input images or limited information about shifts between LR images provided in the input data. This discussion sets the stage for the motivation behind employing graph neural networks (GNNs) [13] for MISR, which is anticipated to address the identified challenges more effectively.

Guided by the following theses, the subsequent chapters of this dissertation are poised to explore GNNs in the context of MISR, aiming to provide both a detailed theoretical analysis and empirical validations through experiments:

- **Thesis 1:** By representing a set of LR images with sub-pixel shifts as a graph, GNNs are capable of processing this graph to yield super-resolution results that are comparable or superior to those achieved by leading MISR architectures based on convolutional networks.

- **Thesis 2:** GNNs can improve their MISR performance by integrating techniques inspired by existing state-of-the-art MISR models based on CNNs, such as individual feature extraction for each LR image, the employment of attention mechanisms, and dynamic and trainable input registration.

- **Thesis 3:** GNNs can reconstruct a scene from a specific point in time by designating a particular reference image from the input LR image set, with the remaining images serving as supplementary information sources to improve super-resolution accuracy. This methodology is anticipated to mitigate visual inconsistencies in regions of high temporal variability and produce a temporally consistent image.

Chapter 2: Related Work

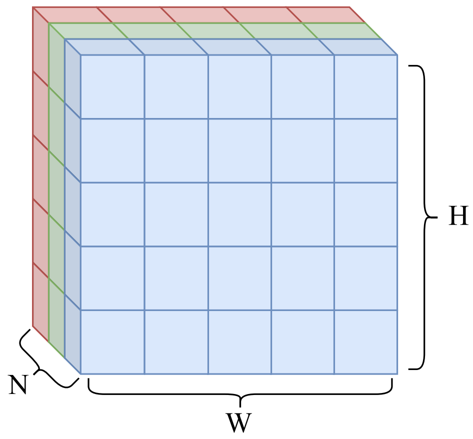
The field of super-resolution reconstruction has seen significant advancements over the years, driven by the increasing demand for high-resolution imagery in diverse domains such as remote sensing, surveillance, and medical imaging. Various methods and techniques have been developed to address the challenges associated with super-resolution reconstruction [1, 4, 6, 15, 18, 20, 21]. In the realm of machine learning, neural networks have played a crucial role in advancing numerous fields, including image and speech recognition, and natural language processing [2, 22]. However, traditional neural networks often struggle with irregularly structured data, as seen in images or time series, and tend to overlook the relational information inherent in the input data. This is where GNNs come into play, showcasing their strength in handling such data.

The chapter then focuses on GNNs, providing a foundation for understanding the advantages they offer in dealing with MISR challenges, addressing traditional limitations of convolutional networks, such as problems with processing irregularly structured data and the tendency to neglect information about relationships present in the input data [17].

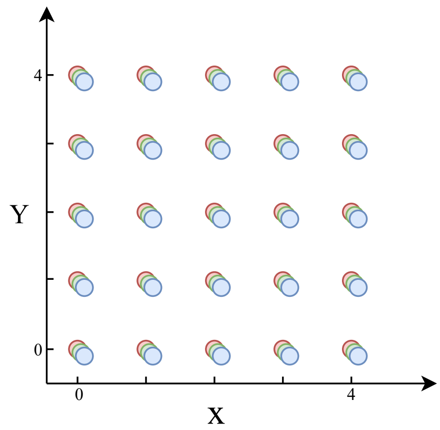
This chapter serves as a comprehensive literature review of MISR techniques, with a particular emphasis on deep learning-based approaches. It outlines the existing research gaps and challenges associated with current MISR techniques, laying the groundwork for the proposed MISR technique using deep GNNs in the following chapter.

Chapter 3: Architecture Design

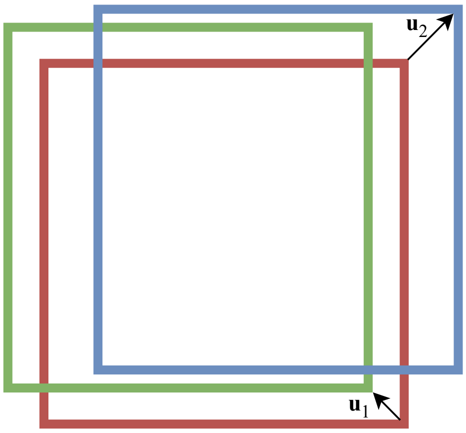
In the initial sections of this chapter, the process of creating a graph from multiple LR images is elaborated. The unique approach amalgamates multiple LR images into a single integrated graph, ensuring a lossless transition while preserving all original information. Moreover, this transition is not only lossless, as it also enriches the volume of information of the input data by specifying providing additional data about sub-pixel translations between



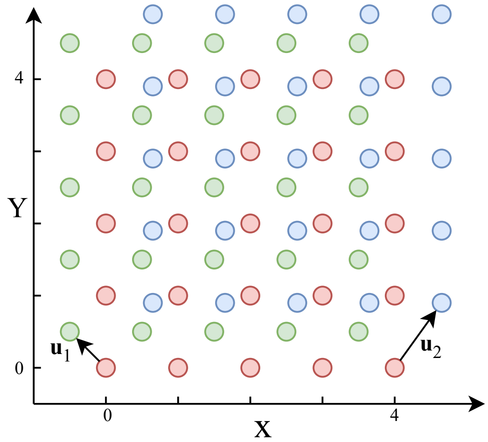
(1) A stack of $N = 3$ input images of dimensions 5×5.



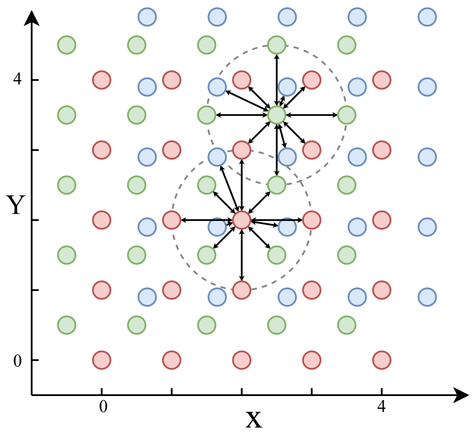
(2) Conversion of pixels to nodes on a mutual 2-dimensional plane.



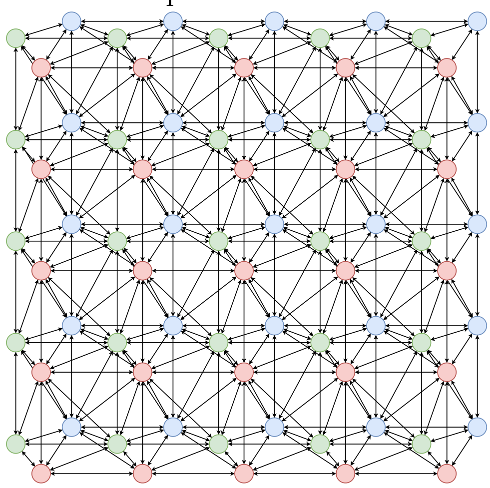
(3) Calculated shift vectors using image registration algorithms.



(4) Node positions adjusted using the computed shift vectors.



(5) Connecting procedure on two arbitrary nodes for radius $r = 1$.



(6) Fully prepared graph.

FIGURE 1: A step-by-step visualization of graph creation procedure.

the input images. This conversion, depicted in Figure 1 is a three-step process:

1. **Node positioning**: The step describes the process of converting pixels from LR images to nodes on a unified plane, with each node's position reflecting the original pixel position, and assigning input features to nodes based on the image channels.

2. **Displacement calculation**: The displacement vectors are determined in this step to align LR images concerning a reference image, aiding in the accurate positioning of nodes in the graph to reflect the spatial relationships and deviations between different LR images.

3. **Graph construction**: In this part, edges between nodes are established based on proximity, using a specified radius, and edge attributes are introduced to capture spatial offsets between nodes, forming a structured graph representation of the original LR images.

The chapter also introduces several models developed to substantiate the second thesis of this dissertation. The models created in this pursuit are discussed in detail, each representing a distinct approach or enhancement towards leveraging the aforementioned techniques for MISR.

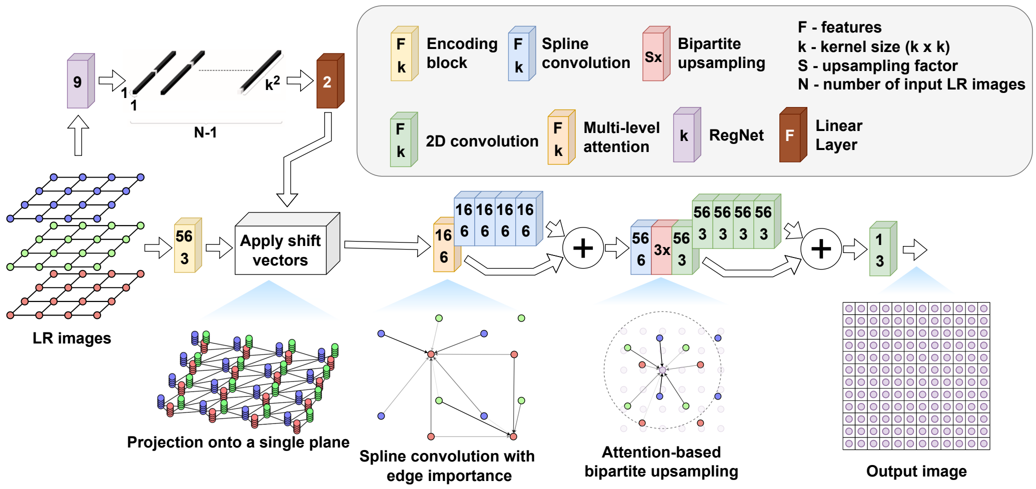


FIGURE 2: The architecture of MagNAt.

- **MagNet**: By processing the input graph, composed of multiple LR images, MagNet is the first GNN used in MISR, yet considered as a proof-of-concept.

- **MagNet++**: This model introduces graph-based upsampling in MagNet, potentially overcoming the limitations of MagNet.

- **$\mathbf{MagNet_{enc}}$**: An improved feature extraction procedure is presented in this model.
- **MagNAt**: This model introduces learnable relationships with attention-based convolution and dynamic registration. It is used as the main GNN model throughout the experimental validation part of this dissertation. The architecture of MagNAt is depicted in Figure 2.
- **$\mathbf{MagNAt_{no\_reg}}$**: This is a modification of the MagNAt model lacking the dynamic registration component, to assess its importance on the super-resolution performance.
- **$\mathbf{MagNAt_{lead}}$**: A model aiming to substantiate the third thesis aiming to reconstruct a scene at a specific point in time dictated by the *leading* LR image.

Chapter 4: Data Description and Simulation

This chapter outlines the datasets employed for the research, underscoring their role in training and validating the proposed models.

- **Simulated Dataset**: Two simulated datasets, SRRB and SRRB_{enh}, of different levels of complexities are created for training and evaluation purposes, providing a controlled environment to test the models. The process of generating these datasets and their significance in the research is elaborated.
- **Real-World Dataset**: The real-world dataset, specifically the Proba-V MISR dataset, is introduced. It discusses the structure of the dataset, its spectral bands, and how real-world challenges are handled. The importance of the Proba-V MISR dataset in the research is also highlighted.

The chapter sets the foundation for experimental validation in subsequent chapters by ensuring a thorough understanding of the data dynamics in MISR.

Chapter 5: Training Methodology and Evaluation Metrics

This chapter delineates the training methodology and evaluation metrics crucial for optimizing and assessing the models' performance in super-resolution tasks. The training hyperparameters for each model, namely HighRes-Net, RAMS, PIUNET, and TR-MISR, are provided, with a note that their specific hyperparameters were derived directly from their corresponding papers, ensuring a faithful reproduction of each model's performance.

- **Training Regimen**: The training protocols adopted for all models are described, underscoring the significance of validation-based performance tuning to prevent overfitting and ensure robust learning.
- **Evaluation Metrics**: The evaluation metrics, namely cPSNR, SSIM, LPIPS, MGE, and TBE, are discussed, providing a comprehensive means to assess and compare the quality of super-resolved images produced by different models. Also, the process of their adoption for addressing MISR-specific challenges is delineated.

This discussion sets the stage for the experimental validation carried out in the subsequent chapter, ensuring a thorough understanding of the methodologies that underpin the research.

Chapter 6: Experimental Results and Discussion

This chapter delves into the comprehensive evaluation of the models developed in this dissertation, with the main focus on MagNAt. It encompasses a range of evaluations on both simulated and real-world datasets, providing critical insights into the capabilities and limitations of the proposed model and comparing it with current state-of-the-art MISR models. This chapter explores the performance of the models from various angles, including their effectiveness in handling simulated scenarios, real-world challenges, temporal variations, architectural progression, and computational efficiency. The following sections provide a succinct overview of the key topics covered in this chapter:

1. **Simulated Datasets**: The models are assessed on simulated datasets through both quantitative (Table 1) and qualitative (Figure 3) analysis, supported by statistical tests. The performance of MagNAt is particularly highlighted, demonstrating its effectiveness in generating high-quality super-resolved images and showcasing its superior handling of simulated scenarios compared to other models.

TABLE 1: Aggregated performance metrics for super-resolution models, combining results from both simulated datasets. The best scores are highlighted in bold, and the second-best scores are underlined.

Dataset	Model	cPSNR	cSSIM	cLPIPS	cMGE	TBE
SRRB	Bicubic	24.56	0.783	0.355	0.129	0.405
	HighRes-Net	29.60	0.913	0.057	0.036	<u>0.314</u>
	RAMS	<u>31.96</u>	<u>0.945</u>	**0.038**	<u>0.021</u>	0.315
	PIUNET	31.49	0.941	0.052	0.026	0.323
	TR-MISR	30.38	0.920	0.055	0.031	0.318
	MagNAt	**32.81**	**0.948**	<u>0.041</u>	**0.019**	**0.310**
$SRRB_{enh}$	Bicubic	24.35	0.749	0.386	0.136	0.396
	HighRes-Net	27.83	0.863	0.118	0.051	0.334
	RAMS	<u>29.10</u>	<u>0.888</u>	<u>0.100</u>	<u>0.037</u>	<u>0.333</u>
	PIUNET	28.90	0.884	0.103	0.038	0.340
	TR-MISR	28.06	0.867	0.115	0.045	0.334
	MagNAt	**30.12**	**0.901**	**0.094**	**0.026**	**0.330**

TABLE 2: Performance metrics obtained by all tested methods on the Proba-V dataset. The best scores are highlighted in bold, while the second-best scores are underlined for each spectral band independently.

Band	Model	cPSNR	cSSIM	cLPIPS	cMGE	TBE
NIR	Bicubic	33.380	.8625	.2791	.0119	.4595
	HighRes-Net	35.401	.9117	<u>.1361</u>	.0065	.3338
	RAMS	35.648	.9148	.1571	.0065	.3382
	PIUNET	35.769	.9127	.1683	.0067	.3510
	TR-MISR	<u>35.958</u>	**.9166**	**.1307**	<u>.0062</u>	<u>.3337</u>
	MagNAt	**36.169**	<u>.9161</u>	.1777	**.0059**	**.3280**
RED	Bicubic	36.419	.9000	.3028	.0068	.4481
	HighRes-Net	37.743	.9337	<u>.1393</u>	.0037	.3245
	RAMS	38.492	<u>.9411</u>	.1601	<u>.0033</u>	<u>.3206</u>
	PIUNET	38.629	**.9430**	.1706	.0035	.3462
	TR-MISR	<u>38.650</u>	.9396	**.1299**	<u>.0033</u>	.3238
	MagNAt	**38.819**	.9406	.1649	**.0032**	**.3195**

2. **Real-World Evaluation**: In this section, the real-world performance of MagNAt on the Proba-V dataset is explored through both quantitative and qualitative analysis, supported by statistical tests, showcased in Table 2 and Figure 4. MagNAt exhibits competitive performance when compared to other models, effectively addressing real-world challenges. The dynamics of performance with varying numbers of input images are also discussed, along with a qualitative analysis on NIR and RED subsets.

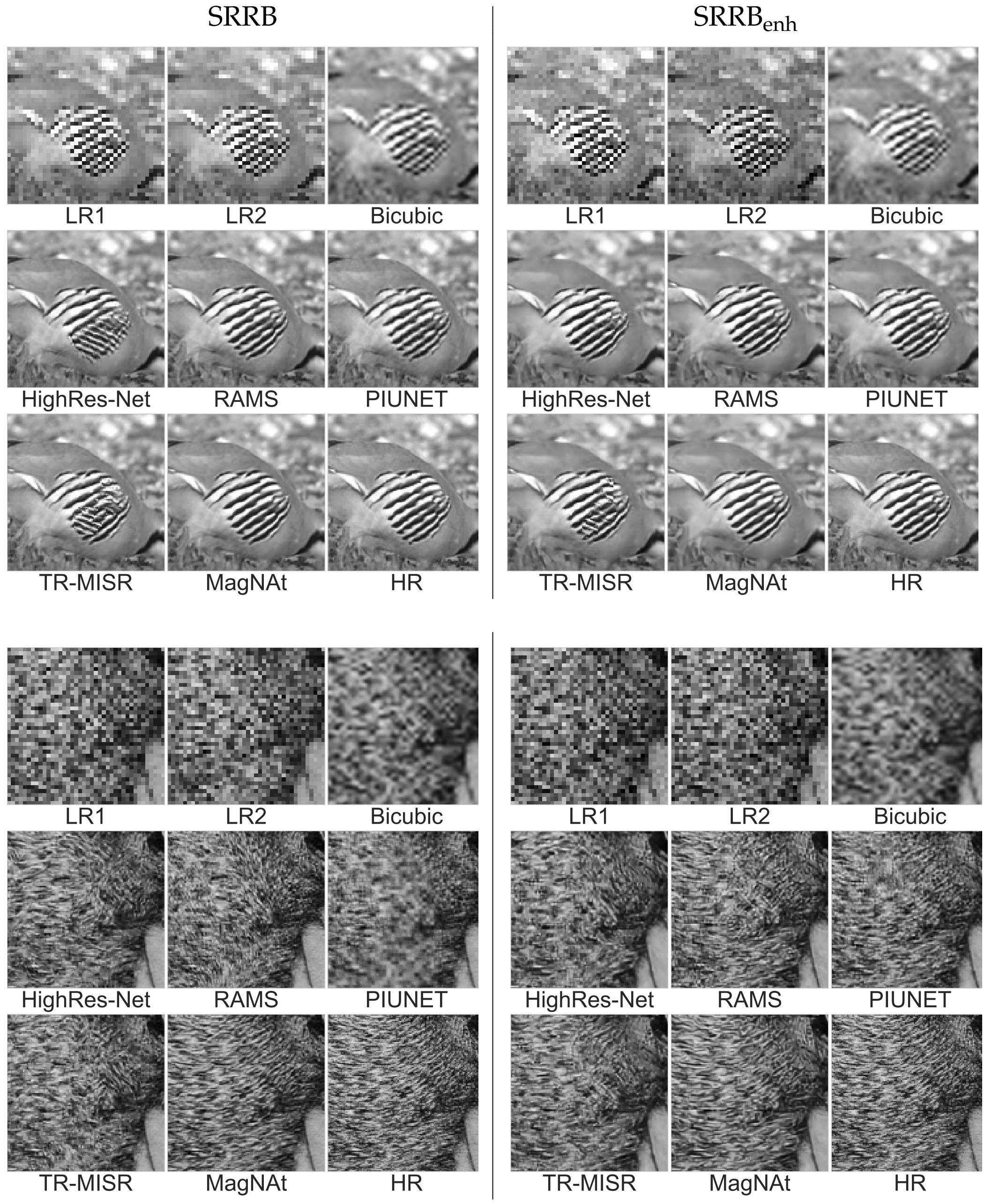
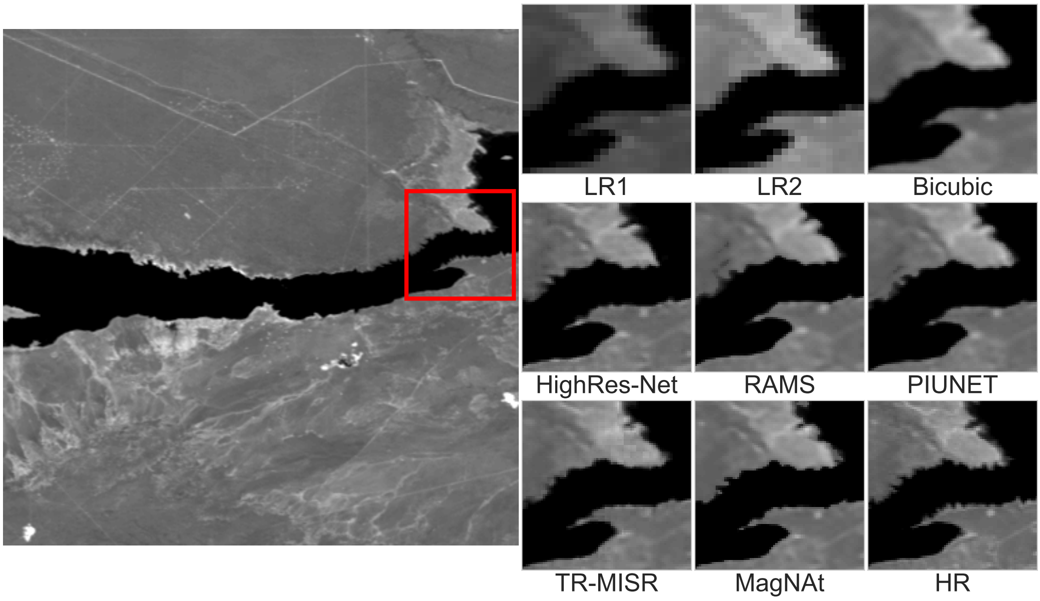


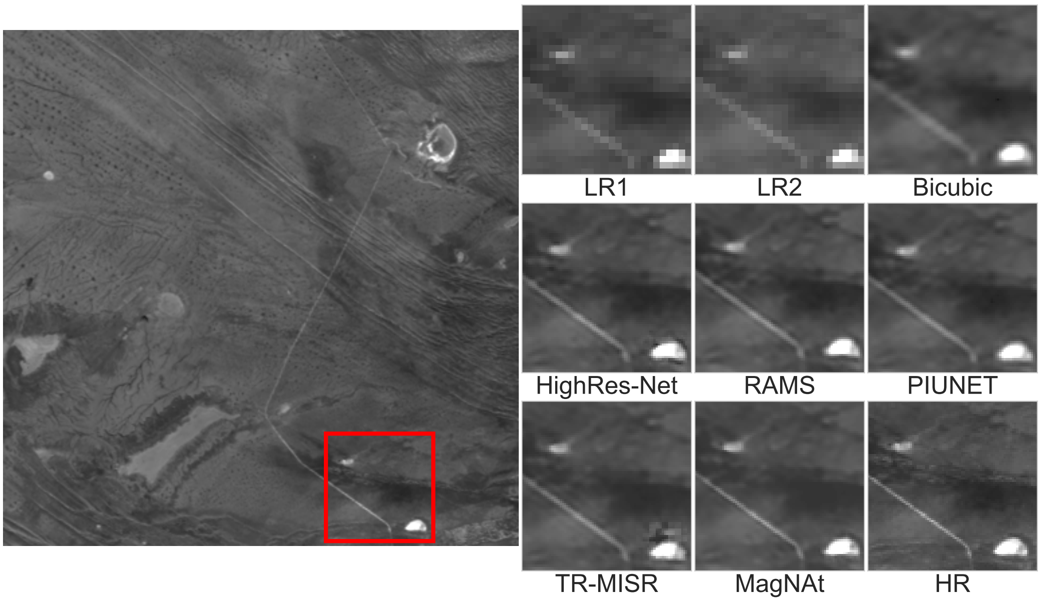
FIGURE 3: Examples of super-resolved simulated images from BSDS100 (top) and Set14 (bottom) datasets.

3. **Temporal Variations and Super-Resolution**: The challenge of temporal variability in input data is discussed, with a focus on how MagNAt mitigates this challenge by guiding the reconstruction process to a specific timeframe dictated by the leading input image. The effectiveness of this approach in producing temporally consistent super-resolved images is highlighted, as shown in Figure 5.

4. **Comparative Analysis of Architectural Progression**: This section discusses the evolution of models developed throughout the dissertation,



(1) Comparison on NIR subset.



(2) Comparison on RED subset.

FIGURE 4: Visual comparisons of different models on the NIR (A) and RED (B) subsets of the Proba-V dataset. The images are cropped to a 150×150 region centred at the same location for all models.

with a focus on how MagNAt, with its attention-based convolution

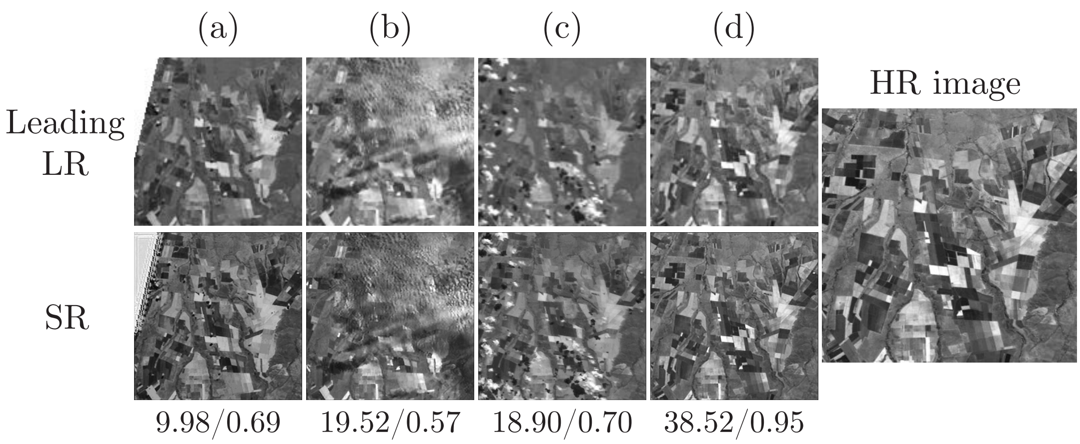


FIGURE 5: Reconstruction results from $\text{MagNAt}_{\text{lead}}$ for the same scene using different leading LR images, with one of them (d) captured at the same time as the HR image. Corresponding cPSNR and cSSIM scores are provided below each super-resolved image.

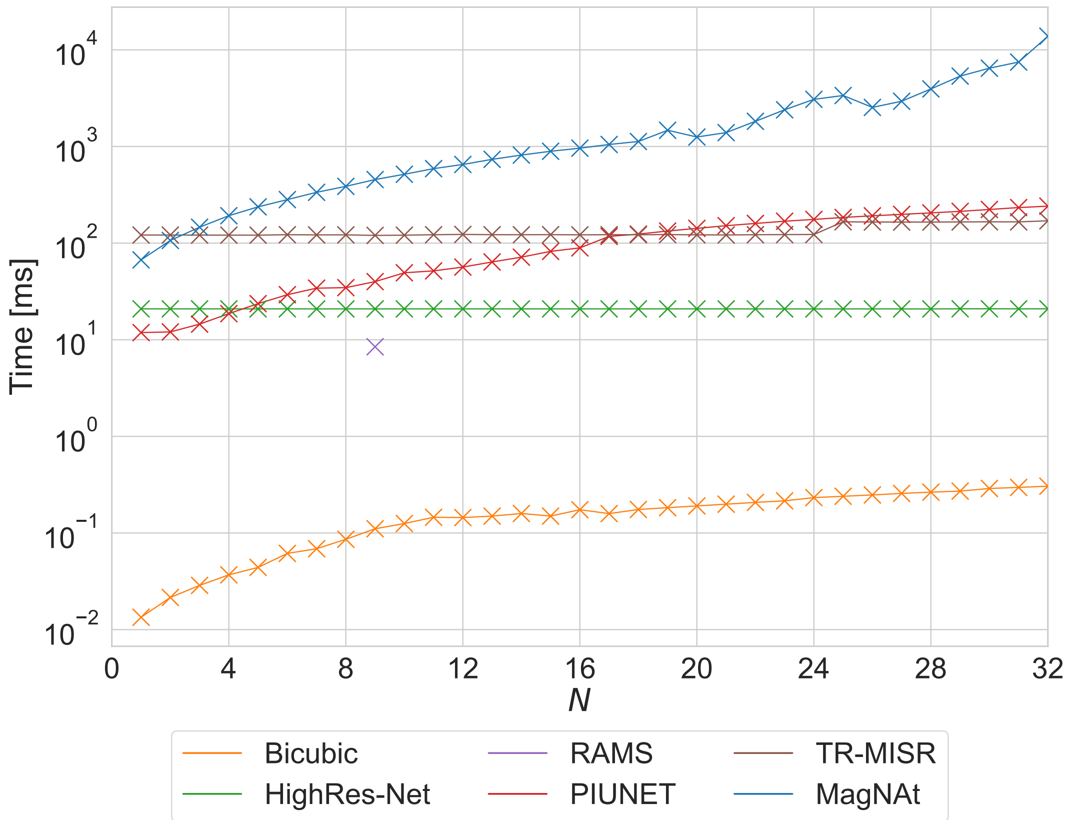


FIGURE 6: Computational times (in milliseconds) against the number of LR images (N). The y-axis is presented in a logarithmic scale.

and dynamic registration, improved MISR performance by integrating techniques inspired by existing state-of-the-art MISR models based on

CNNs.

5. **Time and Memory Analysis**: A detailed analysis of the time and memory requirements of MagNAt, shown in Figure 6 revealed its computational demands. While the model shows promise in achieving high-quality super-resolution results, the analysis identified opportunities for optimization to improve its efficiency and make it more suitable for real-time applications.

Chapter 7: Summary and Conclusions

The dissertation discussed the significance of MISR, set the context for GNNs in MISR, and established the research's theses. A comprehensive literature review highlighted gaps and challenges in existing MISR techniques, paving the way for the proposed GNN-based approach. The creation of a graph from LR images and the introduction of GNN-based models were discussed, along with their enhancements. Simulated and real-world datasets were chosen for evaluation. The chapter provided training details for models and introduced evaluation metrics, setting the stage for rigorous experimentation. The research presented results and discussions on simulated and real-world datasets, validated the theses, and addressed temporal variations and model enhancements. A detailed analysis of time and memory requirements shed light on computational efficiency and scalability, with optimization opportunities identified.

The future directions were also discussed, including a focus on optimizing computational efficiency and scalability, exploring more efficient graph construction methods, and leveraging advanced hardware for faster computations. Also, adapting the model for multispectral MISR represents a promising research avenue. Moreover, exploring handling rotations, non-rigid transformations, and spatially irregular data processing can extend the model's applicability to complex MISR scenarios.

Bibliography

[1] Tai An et al. "TR-MISR: Multiimage Super-Resolution Based on Feature Fusion With Transformers". In: *IEEE J-STARS* 15 (2022), pp. 1373–1388.

[2] Liqiang Bao et al. "Masked Graph Attention Network for Person Re-Identification". In: *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. 2019, pp. 1496–1505. DOI: 10.1109/CVPRW.2019.00191.

[3] Alan C. Bovik. *Handbook of Image and Video Processing (Communications, Networking and Multimedia)*. USA: Academic Press, Inc., 2005. ISBN: 0121197921.

[4] Michel Deudon, Alfredo Kalaitzis, et al. "HighRes-net: Recursive Fusion for Multi-Frame Super-Resolution of Satellite Imagery". In: *arXiv preprint arXiv:2002.06460* (2020).

[5] Kunio Doi. "Computer-aided diagnosis in medical imaging: Historical review, current status and future potential". In: *Computerized Medical Imaging and Graphics* 31.4 (2007). Computer-aided Diagnosis (CAD) and Image-guided Decision Support, pp. 198–211. ISSN: 0895-6111. DOI: https://doi.org/10.1016/j.compmedimag.2007.02.002. URL: https://www.sciencedirect.com/science/article/pii/S0895611107000262.

[6] Chao Dong, Chen Change Loy, and Xiaoou Tang. "Accelerating the Super-Resolution Convolutional Neural Network". In: *CoRR* abs/1608.00367 (2016). arXiv: 1608.00367. URL: http://arxiv.org/abs/1608.00367.

[7] Sina Farsiu et al. "Fast and robust multiframe super resolution". In: *IEEE Trans. on Image Process.* 13.10 (2004), pp. 1327–1344.

[8] Keith Fife, Abbas El Gamal, and H.-S. Philip Wong. "A 3MPixel Multi-Aperture Image Sensor with 0.7m Pixels in 0.11m CMOS". In: *2008 IEEE International Solid-State Circuits Conference - Digest of Technical Papers*. 2008, pp. 48–594. DOI: 10.1109/ISSCC.2008.4523050.

[9] Rafael C. Gonzalez and Richard E. Woods. *Digital image processing*. Upper Saddle River, N.J.: Prentice Hall, 2008. ISBN: 9780131687288 013168728X 9780135052679 013505267X. URL: http://www.amazon.com/Digital-Image-Processing-3rd-Edition/dp/013168728X.

[10] J.R. Jensen. *Remote Sensing of the Environment: An Earth Resource Perspective*. Prentice Hall series in geographic information science. Pearson Prentice Hall, 2007. ISBN: 9780131889507. URL: https://books.google.pl/books?id=A6YsAQAAMAAJ.

[11] Rajandeep Kaur and. "A Review of Image Compression Techniques". In: *International Journal of Computer Applications* 142 (May 2016), pp. 8–11. DOI: 10.5120/ijca2016909658.

[12] Michal Kawulok et al. "Deep learning for multiple-image super-resolution". In: *IEEE GRSL* 17.6 (2020), pp. 1062–1066.

[13] Maciej Krzywda, Szymon Lukasik, and Amir H. Gandomi. "Graph Neural Networks in Computer Vision - Architectures, Datasets and Common Approaches". In: *2022 International Joint Conference on Neural Networks (IJCNN)*. IEEE, July 2022. DOI: 10.1109/ijcnn55064.2022.9892658. URL: https://doi.org/10.1109%2Fijcnn55064.2022.9892658.

[14] Y. Lecun et al. "Gradient-based learning applied to document recognition". In: *Proceedings of the IEEE* 86.11 (1998), pp. 2278–2324. DOI: 10.1109/5.726791.

[15] Andrea Bordone Molini et al. "DeepSUM: Deep neural network for super-resolution of unregistered multitemporal images". In: *IEEE TGRS* 58.5 (2020), pp. 3644–3656.

[16] Kamal Nasrollahi and Thomas Moeslund. "Super-resolution: A comprehensive survey". In: *Machine Vision and Applications* 25 (Aug. 2014), pp. 1423–1468. DOI: 10.1007/s00138-014-0623-4.

[17] Kamal Nasrollahi and Thomas B Moeslund. "Super-resolution: a comprehensive survey". In: *Machine vision and applications* 25.6 (2014), pp. 1423–1468.

[18] Francesco Salvetti et al. "Multi-Image Super Resolution of Remotely Sensed Images Using Residual Attention Deep Neural Networks". In: *Remote Sensing* 12.14 (2020), p. 2207.

[19] Yunqing Tang, Yin Xiang, and Guangfeng Chen. "A Nighttime and Daytime Single-Image Dehazing Method". In: *Applied Sciences* 13.1 (2023). ISSN: 2076-3417. DOI: 10.3390/app13010255. URL: https://www.mdpi.com/2076-3417/13/1/255.

[20] Tomasz Tarasiewicz et al. "Multitemporal and multispectral data fusion for super-resolution of Sentinel-2 images". In: *IEEE Transactions on Geoscience and Remote Sensing (TGRS)* (in press).

[21] Diego Valsesia and Enrico Magli. "Permutation Invariance and Uncertainty in Multitemporal Image Super-Resolution". In: *IEEE TGRS* 60 (2022), pp. 1–12.

[22] Sijie Yan, Yuanjun Xiong, and Dahua Lin. *Spatial Temporal Graph Convolutional Networks for Skeleton-Based Action Recognition*. 2018. arXiv: `1801.07455 [cs.CV]`.