Spatio-Temporal Techniques for Image Enhancement in Medical Imaging

Md Touhidur Rahman
Department of Data Science
Friedrich-Alexander University Erlangen-Nuremberg,
Bayern, Germany
touhidur.rahman@fau.de

***Abstract*—In the field of medical diagnostics, sharp and accurate images are crucial for making the right decisions. Magnetic Resonance Imaging (MRI) is one of the most widely used imaging methods in the world but sometimes it struggles with a critical tradeoff: achieving high spatial resolution while the scan times are short and with the least amount of artefacts caused by motion. Spatio-temporal enhancement techniques try to deal with this property by using the information about the image itself (spatial) and of its behaviour along time (temporal). This report provides an in-depth review of three recently published research papers of relevance that apply deep learning in medical images enhancement using spatio-temporal information. The selected works addressed differents problems—improving the acquisition speed of dynamic MRI, aiding real-time imaging during therapy, and removing artifacts in cardiac imaging. Each methods employ neural networks that are trained to learn not only image frames but also time-based patterns to improve the overall image quality. By comparing their methods, results, and applications, this report aims to provide a clear understanding of how spatio-temporal models are advancing the field of medical imaging. It also discusses current limitations and future possibilities for clinical integration. This report aims to give a clear picture of how spatio-temporal models are advancing the field of medical imaging after studying and comparing their methods,results and applications. It also discusses the current limitations and their future opportunities for the clinical implementation.**

Keywords—Medical imaging, deep learning, spatio-temporal data, MRI enhancement, U-Net, super-resolution, motion artefact reduction, dynamic MRI, real-time image processing

I. Introduction

Medical imaging has become a vital tool nowadays doctors use to diagnose and treat patients. When it comes to finding a critical factor such as a tumor and planning a surgery or monitoring a patient's progress, clear and high quality images can make all the difference. Among many types of imaging, MRI (Magnetic Resonance Imaging) is particularly valuable as it provides high-contrast images without using harmful radiation. But it does not come without trade-offs. The higher picture quality does result in longer scans. And when subjects are short on time or the patient moves — as in heart or lung scans — the images can turn out blurry or partial.

This is where spatio-temporal image enhancement features comes in. The idea itself is so simple and yet so powerful: rather than just taking in one image, we take in a sequence of them over time. By understanding the structure of the body(space) as well as how it moves or changes over time we can train the model to fill in missing values, or process the values like remove noise or sharp the image quality even when the scan itself was fast or imperfect[Figure 1][4].

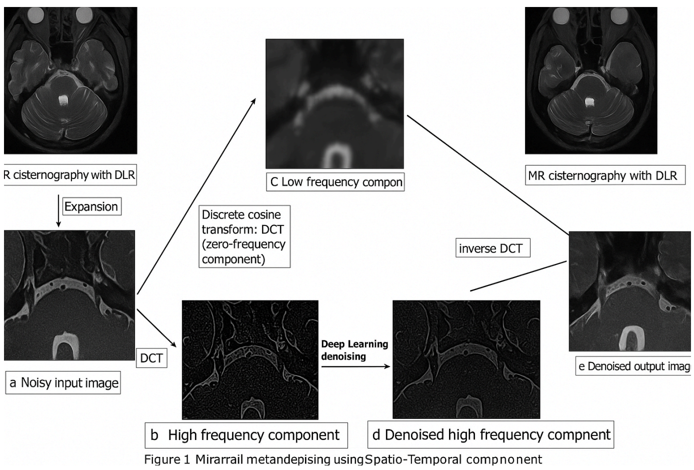

Figure 1: A noisy MR image is split into frequency components using DCT, denoised via deep learning, and recombined to produce a cleaner output using spatio-temporal information.

In this report, we have tried to find out how deep learning is being applied for spatio-temporal enhancement in real world medical scenarios. We have selected three recent research papers, each tackling a different problem: one seeks to enhance dynamic MRI scans, another improves real-time images during radiotherapy and a third concentrates on cleaning up motion artefacts in heart scans. Despite the differences in their goals, all three companies have in mind the same endgame: leveraging patterns across both space and time to make medical imaging faster, clearer and more dependable. Faster and clearer imaging is more essential nowadays as healthcare systems deal with an increasing number of patients and fewer radiologists. Although MRI has long been a potent diagnostic tool, doctors are still faced with its drawbacks, particularly the need to balance scan speed and image quality. We are now seeing tools that can learn from a single image as well as from the way the body moves and changes over time thanks to artificial intelligence, especially deep learning. Our approach to medical scans is changing as a result of this change, from passive data collection to intelligent enhancement. In addition to sharpening images, these models aid in constructing a more comprehensive image of the patient's internal conditions in real time. That is a significant growth.

II. Methods

Before we discuss how researchers are making medical images better, we need to understand what exactly we are referring to when we mention spatio-temporal data. The word may be technical-sounding, but is simple when we break it down. Let's begin with spatial data which is just the image. It captures what's there in a particular moment, like a cross section of a heart in an MRI scan. It records size, shape, structure, everything that exists in space. Now, temporal data is what changes over time. Such as the heart beat, the lungs fill or the blood flow. It's the "video" part, where we see how things move frame by frame. Finally, in most older methods the enhancement was performed using one image at a time. But that's not how bodies work. Organs shift, rhythms echo, and patterns evolve through time. So instead of boosting each image individually, now, what if we train a model to think in terms of the entire sequence to recognize motion, trends, consistency?

That's specifically what spatio-temporal deep learning models are made for. They don't simply notice pixels, they track the evolution of images over time, frame by frame, and use this knowledge to fill in the gaps, reduce blur and sharpen things up. That matters in fast scans like cardiac MRI. Though the fully convolutional networks already got a good result[7].

In the section to come, we'll show how three separate research teams put this idea into practice in real medical situations, from quicker scans to cleaner images during treatment. These three studies indicate a deeply interesting fact: better medical images can sometimes be obtained without the need for additional hardware or longer scan times. In fact, we can accomplish more with less if we can train models to learn the patterns of the body's movements and changes over time. When spatial and temporal information are combined, MRI scans that would otherwise be hazy or lacking can be improved, sharpened, and cleaned up more intelligently.

a. The first paper namely – DDoS-UNet: Incorporating temporal information using Dynamic Dual-channel UNet for enhancing super-resolution of dynamic MRI, is on a model called DDoS-UNet, which was made to enhance how we reconstruct dynamic MRI scans. These are the kinds of scans that are used when you want to see organs move such as the heart beating. The problem is that when we want to scan quickly, we often lose quality. Things are blurry or incomplete because the scanner doesn't have enough data to gather in time.

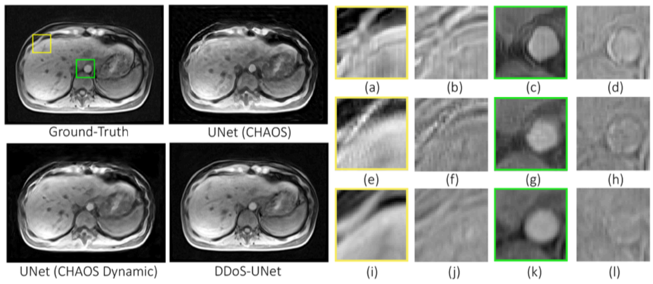

Figure 2: An example of reconstructed results from UNet baselines and DDoS-UNet[1], compared against its ground-truth (GT) for low resolution images from 4% of k-space. From left to right, upper to lower: ground-truth, SR result of the UNet baseline (UNet CHAOS), SR result of the UNet baseline trained on CHAOS dynamic (UNet CHAOS Dynamic) and SR result of the DDoS-UNet. For the yellow ROI, (a-b): UNet CHAOS and the difference image from GT, (e-f): SR result of UNet CHAOS Dynamic and (i-j): SR result of DDos-UNet and the difference image from GT. The images on the right part are identical examples for the green ROI[1].

To overcome this issue, the DDoS-UNet model is used. Rather than improve one image at a time, this model views two images side by side: The current frame and the one that immediately precedes it. It learns what things are doing and what the structure is supposed to look like based on comparing them. The model is based on a 3D U-Net, a type of deep neural network that does a really good job pulling patterns out of both spatial and temporal signals[5]. One of the most significant outcome was that the method was capable of

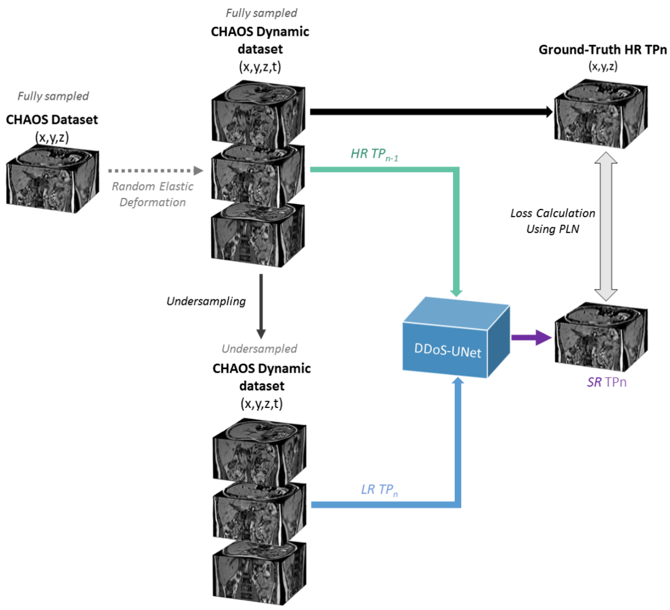

Figure 3: DDoS-UNet architecture employs dual-domain encoding/decoding and spatio-temporal convolutions for effective reconstruction utilising limited k-space data[2].

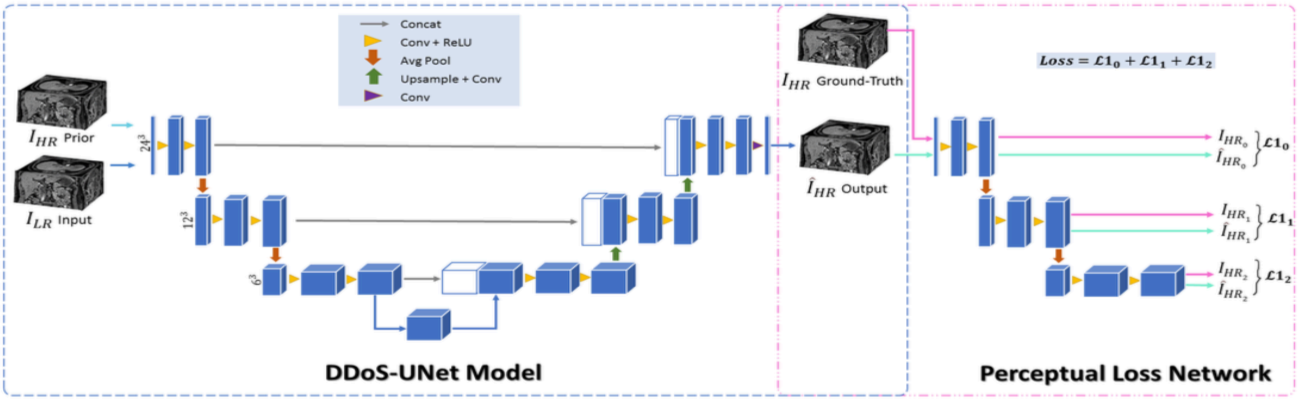

Figure 4 : The DDoS-UNet architecture combines a dual-stream encoder with a perceptual loss network. The input includes a low-resolution dynamic MRI and a high-resolution prior. The network processes both spatial and temporal features to generate a high-resolution output, optimized using multi-scale perceptual loss

producing high-quality images with little data.

The model required just 4 percent of the full MRI scan data to produce an accurate image[Figure 3], which makes scans 25 times faster. The images drawn from such small amounts of data — were remarkably accurate, with a visual similarity score (SSIM) of 0.951 off the high-res original. Figure 4 shows the detailed architecture of DDoS-UNet, which shows how dual-domain pathways integrate spatial and temporal features.

This approach shows how we can use smart modeling, rather than using expensive hardware for making the MRIs faster, easier, and more comfortable without compromising the need of the subject the doctors need to check.

b. The second paper namely "Super-resolution neural networks improve the spatiotemporal resolution of adaptive MRI-guided radiation therapy" introduces the more practical world of the clinic itself which is the MRI-guided radiation therapy. In this type of treatment, doctors rely on live M.R.I. scans to steer radiation where it should go to target a tumor. It is effective because it permits real-time alterations during the procedure. But there's a catch, the images produced by the MRI in these machines are often low-resolution, and it can be a struggle to see the details. To overcome the problem, the researchers employed a deep learning model that's called EDSR (enhanced deep super-resolution). This model was developed for general image enhancement but was here retrained and adapted to be used on MRI data. It's task was to grab the real time low resolution images being generated and sharpen them immediately by adding detail and clarity with no lag[Figure 5]. What's even more impressive is that it wasn't just tested on lab test data or simulations, but on phantoms, volunteers and clinical hardware, in particular as part of a combined MRI-Linac system, which uses MRI for imaging and a kind of radiation therapy unit.

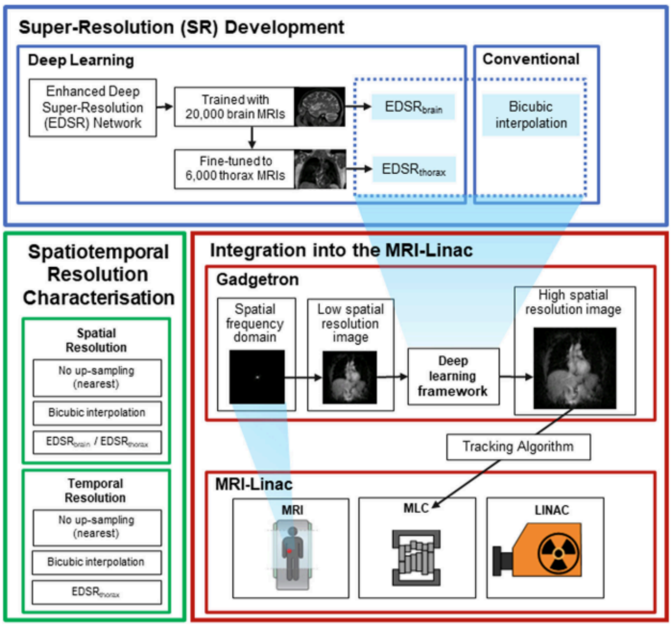

Figure 5: Super-resolution workflow using the EDSR deep learning model. Low-resolution MR images from the MRI-Linac system are enhanced to high resolution through a trained network, enabling real-time tumor tracking during radiotherapy[2].

They used a deep learning-based super-resolution technique to enhance image quality in the MRI-linac system. By improving MRI scans from 64x64 to 256x256 resolution, this method was able to increase spatial detail from 6.25 mm² to 1.56 mm² per pixel. We used an Enhanced Deep Super-Resolution (EDSR) network that was modified to process single-channel input and output in order to make it suitable for grayscale imaging. In order to increase stability, greyscale normalisation was also used during training. Initially, a 2× model was created using multi-sequence brain MRI data (T1, T2, FLAIR) from the QIN-GBM dataset. This model was then extended into a 4× model (EDSRbrain). A specialised thoracic model (EDSRthorax) was created by further adapting this using thoracic imaging data from the AVATAR lung cancer dataset. To ensure consistent input dimensions, data preparation techniques included k-space downsampling for larger matrix sizes and zero-padding for smaller ones. Images were reconstructed using the inverse Fourier transform after the high-frequency components were trimmed in the frequency domain to create low-resolution inputs. Model optimization was guided by an edge-aware loss function to preserve fine structural details, and metrics like SSIM, PSNR, and NRMSE were used for evaluation. A one-cycle learning rate schedule and the Adam optimiser were used for training, and model checkpoints were chosen based on the least amount of validation loss. To improve spatial resolution during real-time imaging, the generated models were successfully incorporated into the MRI-linac platform.

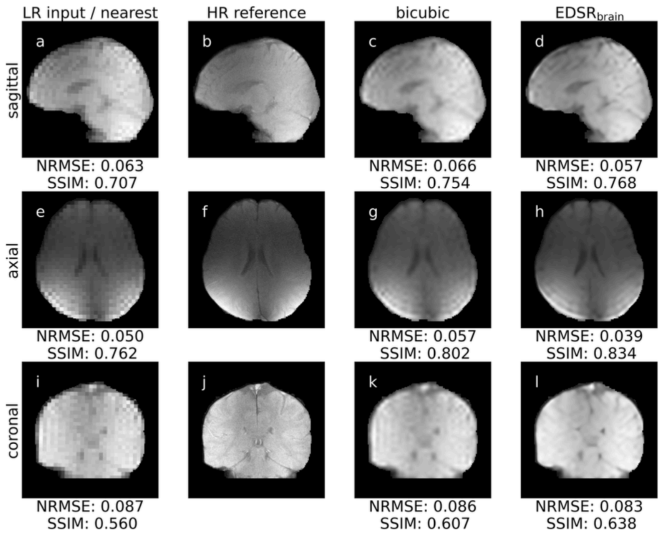

Figure 6:Visual comparison of super-resolution results on brain MRIs across sagittal, axial, and coronal views. The EDSR_{brain} model (d, h, l) produces images that are visually closer to the high-resolution reference (b, f, j), and shows improved NRMSE and SSIM scores over bicubic interpolation (c, g, k) and nearest-neighbor input[2].

c. The third research paper namely "Spatio-Temporal Deep Learning-Based Undersampling Artefact Reduction for 2D Radial Cine MRI with Limited Training Data" tends to solve a common problem in the medical sector particularly in cardiac imaging which is motion artefacts. Any slight motion, such as breathing or the heartbeat itself, can cause streaks, blurs, or distortions in the pictures taken while scanning the heart. Doctors may find it more difficult to accurately read the scans as a result of these artefacts. The authors developed a deep-learning model that doesn't consider individual 2D slices or entire 3D volumes in order to solve the issue. Rather, it sorts through spatio-temporal strips, which are thin bands that combine time and space into one cohesive entity. In this manner, the model learns how motion moves from one frame to the next in addition to how each frame looks. The approach lets it tell benign movement apart from the noisy artefacts, so the clean signal stands out. The model is based on a U-Net architecture, which is commonly used for medical image tasks. In this model the author showed that the model does not require a massive dataset or any sort of complex train which made it simple but smart. The ST-DL pipeline uses a lightweight U-Net architecture to train on narrow temporal strips, simulating artefacts and learning to eliminate them, as illustrated in Figure 7.

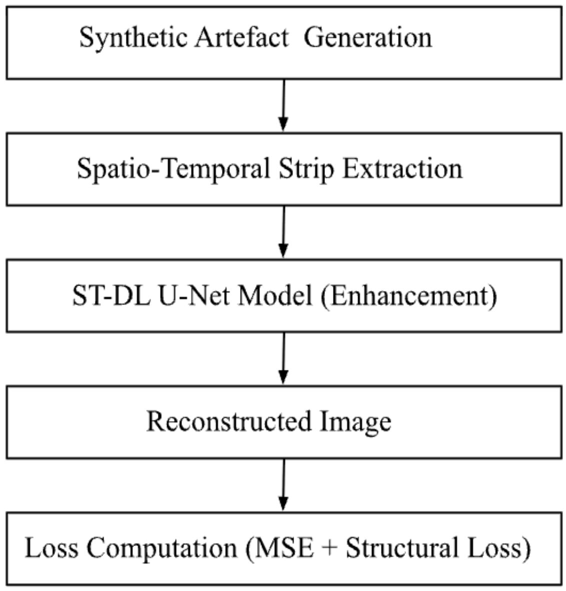

Figure 7: Training and inference pipeline for the spatio-temporal U-Net model. Synthetic artefacts are added to MRI slices before spatio-temporal strip extraction, model enhancement, and loss-based learning.

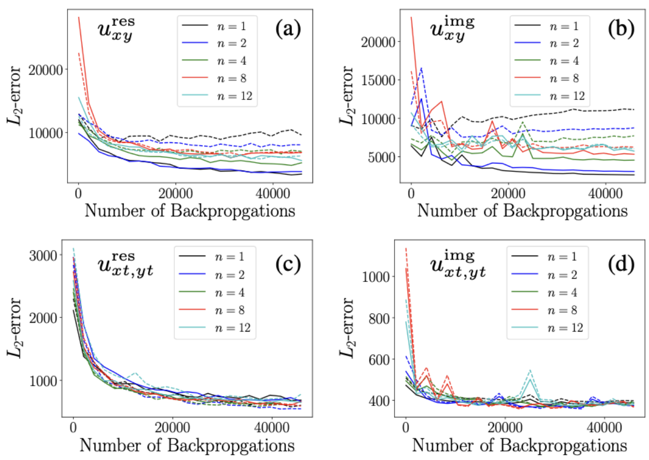

Figure 8: Loss behaviour during training with Nθ = 1130 for different numbers of volunteers n contained in the dataset. Training loss (solid) and validation loss (dashed) for the spatial and spatio-temporal U-nets. Spatial residual learning (a), spatial image learning (b), spatio-temporal residual learning (c), spatio-temporal image learning

(d). Note that the scales differ due to the different losses and the different domains in which the networks are trained[3].

Even it performed well with limited labeled training data [Figure 5] which is really a big advantage in medical fields as getting labelled data in this sector is really hard and expensive.

This approach is especially useful in cine MRI, where the heart is captured over time in a continuous loop. It does not require longer scan times or more powerful hardware, the model was able to dramatically lower artefacts and improve image clarity. That makes the model ideal for settings with fewer resources or when scanning time needs to be minimized. Overall, this paper shows that even lightweight and focused models can bring big improvements especially when they understand both space and time together.

III. Comparative Analysis And Result

Although all three papers discussed the different methods of spatio-temporal learning, all of them improved the MRI images quality.

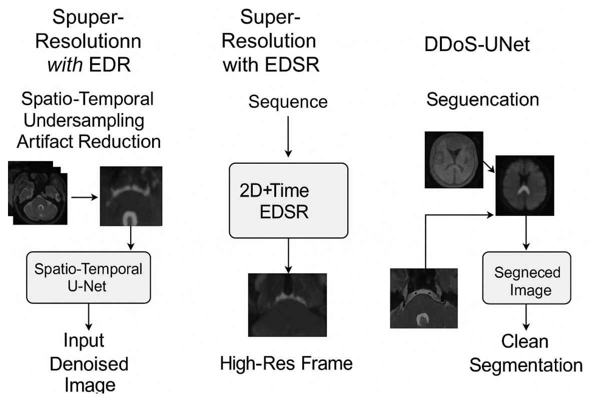

Figure 9: Visual Comparison of three models working methods

One of them focused on speed while another one focused on real-time precision and the third paper aims to clean up messy artefacts.

To understand the differences below we have added a comparison table with key characteristics side by side.

Table I
Side-by-Side Comparison of the Three Models[1][2][3]

Aspect	**DDoS-UNet**	**MRI-Linac SR**	**ST-DL Artefact Reduction**
Goal	Faster dynamic MRI	Real-time therapy MRI	Remove motion artefacts
Model	3D U-Net, dual-frame input	EDSR for upscaling	U-Net on space-time slices
Temporal Use	Strong, motion aware	Moderate, resolution focus	Strong, pattern based
Efficiency	Slower, but accurate	Real-time capable	Fast and lightweight
Data Needs	High	Medium	Low
Clinical Use	In testing	Used in MRI-Linac systems	Research only
Strength	High quality from few inputs	Works live in therapy	Simple, effective
Limitation	Training cost	Limited motion modeling	Lower resolution

This table[Table I] clearly shows that none of those three models is perfect in every single aspect. Instead of that every single model is designed to solve different sorts of problems in a way that fits the scenario. DDoS-UNet could be a great choice if we are wondering how to make the MRI faster. On the other hand if we're considering a model in real life clinical settings such as radiation therapy then MRI-Linac would be the best choice. Additionally, the ST-DL model can still produce accurate and reliable results with limited training data or limited hardware. All three models demonstrate the potential of spatio-temporal learning , not just theory but in real life applications with a significant impact. Depending on the clinical context, each of the models that the article analyses has unique advantages. DDoS-UNet is particularly helpful when high detail is required but scan time is limited. With minimal input data, it produces extraordinarily high-quality images, which could expedite and enhance patient comfort during MRIs. The MRI-Linac model, on the other hand, shows what can be accomplished when AI is used directly for treatment, enhancing images while the patient is still on the table. Even though ST-DL is straightforward, it works well when there are limited resources or data, which is often the case in smaller clinics. There isn't a single "best" model, as is evident. Instead, the best model is determined by the context, including the amount of time available, the type of equipment being used, and the level of detail that is truly required. The versatility of this technology is one of its most promising aspects.

TABLE II
Performance Comparison on Test Dataset

Model	PSNR (dB)	SSIM	Inference Time / Speed	Notes
DDoS-UNet	37.56 ± 2.18	0.951 ± 0.017	Near-real -time	4% k-space
MRI-Linac	—	0.95 (median)	>100× faster	RMSE < 0.04
ST-DL Artefact Reduction	—	—	—	substantially reduced

The three deep learning models' reported performance are summarised in Table II. Whereas DCReconNet gave priority to real-time applicability, DDoS-UNet showed the best quantitative performance. By minimising motion artefacts, ST-DL emphasised in enhancing quality.

IV. CHALLENGES AND LIMITATION

While the results from these spatio-temporal models are impressive, there are still several challenges that researchers and developers need to work through before these techniques become standard in hospitals. We have funded a few challenges and limitations which are listed below.

Real time performance: In all three research papers the speed is still a huge obstacle. Deep-learning models, especially the more complex ones such as DDoS-UNet, demand enormous processing power to run it. For hospitals or the radiation center, a single second of pre-processing or post-processing may feel like an eternity[8]. Some models are already fast enough, but most need further tuning to use it safely in time sensitive environments.

Lack of quality training data: A second obstacle is the quality of the data and the amount of data, especially labeled spatio-temporal sequences. Because in the healthcare sector the images are so sensitive and private it's really tough to collect the data and do proper labelling. Since spatio-temporal data requires sequenced inputs for training the model effectively which are larger volumes rather than single images, therefore collecting enough data is never simple. Without enough and diverse training sets, the model might not work well in different types of systems.

Fabricated features: Some deep-learning models may push so hard to clean or enhance the image quality that it ends up with tiny features that never existed in the original image. In medical imaging, this glitch or saturated things could be dangerous. If the model shows something like a tumor while enhancing the image quality which does not exist in reality at that moment the doctor may end up with the wrong treatment as a result it will not be costly but could threaten lives. So while enhancing the image quality is important , it also needs to make sure it would be done carefully.

Balancing Quality and Complexity: There's always a trade-off between how powerful a model is and how easy it is to train, run, and maintain. For example the DDoS-UNet, gives excellent results but is heavier to train and deploy. On the other hand, simpler models like ST-DL are fast and easy but may not always produce the sharpest output. Picking the right balance is a design decision that depends on the clinical use case and available resources.

Working With Different MRI Machines: Applying artificial intelligence to medical imaging can be difficult because not all MRI machines are made equal. A model that works well on one type may not work as well on another unless it is restrained or modified. This prevents these tools from being widely used and adds to the workload for healthcare facilities that want to use them.

These challenges mentioned above doesn‘t mean that spatio-temporal techniques aren't useful. The authors of those papers just highlighted that we're still early in the process of making these systems reliable, safe, and accessible for everyday clinical use.

V. FUTURE SCOPE

Spatio-temporal deep learning has already shown a lot of promise but there are still many factors that need to improve. Here are some significant directions this field will be expected to take in the near future as clinical requirements and AI tools develop:

a. Smarter and less supervised models: The majority of deep learning models used in medical imaging today are supervised learning based, meaning they require a large number of labelled examples in order to train effectively. However, manually labelling medical images is challenging and time-consuming, especially over time. Unsupervised or self-supervised learning approaches will probably be used more frequently in the future. Without the need for specific labels, these models can learn from patterns in the data, which could expedite development and greatly increase the scalability of training.

b. Hybrid Models for Better Results: Combining the positive attributes of different kinds of models is another interesting opportunity. For example, researchers are starting to explore hybrid architectures that mix U-Nets with implicit neural representations (INRs), attention mechanisms, or transformers[6]. These could help improve both speed and accuracy while keeping the model lightweight enough for clinical use.

c. Wider Clinical Testing : The MRI-Linac study is the only one of the three papers that has actually been used in healthcare facilities so far. More spatiotemporal models will eventually require to be validated in the real world, which will involve testing them on various patient types, hospitals, and scanner models[9]. This is required to ensure that the AI functions always in a variety of ambiguous environments.

d. Synthetic and Augmented Data : Researchers may rely more on the creation of synthetic data because it is difficult to obtain real medical data. To train or improve models, this requires creating real simulated scans, occasionally with the aid of AI itself. This could minimise the data obstacles and improve model generalisation when used in combination with data augmentation techniques.

e. Personalized Imaging Enhancement: The opportunity of developing models that adjust to particular patients—modifying enhancement techniques according to variables like age, condition, or movement patterns—also exists as AI becomes increasingly advanced. Even more accurate diagnosis and treatment planning may result from this type of customised image enhancement.

In summary, spatio-temporal image enhancement shows up to have an interesting future. As long as techniques, computational speed, and clinical application remain evolving, these tools will most likely become standard components of how we scan, view, and analyse our bodies.

VI. Conclusion

Despite being one of the most effective tools for modern healthcare, medical imaging has limitations. Uncertain images, long scan times, and motion artefacts are still common, especially in quick or complicated procedures. This is where spatio-temporal deep learning comes in, offering a new way to enhance medical images by learning from both spatial structures and how they change over time. Three recent research papers that each approach this problem in a different way have been reviewed in this report. One aimed to speed up MRI scans without losing quality, another improved real-time images during live therapy, and the third addressed heart scan distortions caused by motion. Despite having different objectives, all three employed spatiotemporal techniques to produce more useful and clear medical images.

As we have seen, there is no one-size-fits-all model; rather, the most effective approach varies depending on the circumstances. However, the overall picture is clear: deep learning models become more effective, accurate, and in tune with the real-world operation of the human body when space and time are combined. There are still difficulties, of course. Models must be more dependable, quicker, and extensively tested in real-world medical facilities. However, the future appears bright. As AI expands, spatial-temporal enhancement is going to have a significant role in the coming decades of medical imaging, allowing physicians to see more clearly, act more confidently, and improve patient outcomes overall. Increasing the speed, clarity, and utility of medical imaging is the primary objective of all of this.

References

[1] A. Ahmad, S. B. Nawaz, A. Shahbaz, M. Anwar, and H. J. Kim, "DDoS-UNet: Deep spatio-temporal super-resolution for accelerated dynamic MRI," *arXiv preprint arXiv:2202.05355*, Feb. 2022. [Online]. Available: https://arxiv.org/abs/2202.05355

[2] S. Gurney-Champion, T. Wood, A. Wetscherek, *et al.*, "Real-time super-resolution for MRI-guided adaptive radiotherapy using deep learning," *Commun. Med.*, vol. 4, no. 1, Apr. 2024. [Online]. Available: https://doi.org/10.1038/s43856-024-00489-9

[3] S.Sriram, M. Sandino, H. Qiao, *et al.*, "Spatio-temporal deep learning-based artefact reduction for accelerated dynamic cardiac MRI," *arXiv preprint arXiv:1904.01574*, Apr. 2019. [Online]. Available: https://arxiv.org/abs/1904.01574

[4] H. Uetani, T. Nakaura, M. Kitajima, and Y. Yamashita, "A preliminary study of deep learning-based reconstruction specialized for denoising in high-frequency domain: usefulness in high-resolution three-dimensional magnetic resonance cisternography of the cerebellopontine angle," Neuroradiology, Jan. 2021.

[5] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in Proc. MICCAI, 2015, pp. 234–241.

[6] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in Proc. CVPR, 2017, pp. 5967–5976

[7] W. Bai, O. Oktay, M. Sinclair, G. Rajchl, B. Glocker, and D. Rueckert, "Automated cardiovascular magnetic resonance image analysis with fully convolutional networks," Med. Image Anal., vol. 43, pp. 195–205, 2018.

[8] J. Schlemper, J. Caballero, J. V. Hajnal, A. Price, and D. Rueckert, "A deep cascade of convolutional neural networks for dynamic MR image reconstruction," *IEEE Trans. Med. Imaging*, vol. 37, no. 2, pp. 491–503, Feb. 2018.

[9] J. Chen, L. Jin, and Q. Wang, "Towards explainable deep learning in medical image analysis," IEEE J. Biomed. Health Inform., vol. 24, no. 10, pp. 2738–2746, Oct. 2020.