Jami Vihavainen

SAR image quality metrics with respect to Super-Resolution

School of Technology and Innovation
Bachelor of Science
Energy and Information technology

Vaasa 2026

UNIVERSITY OF VAASA	
School of Technology and Innovation	
Author:	Jami Vihavainen
Title of the thesis:	SAR image quality metrics with respect to Super-Resolution
Degree:	Bachelor of Science
Degree Programme:	Automation and Computer Science
Supervisor:	Jouni Lampinen
Year:	2026 **Pages:** 50

ABSTRACT:

Synthetic aperture radar is a type of radar imaging technique, and it can produce high-resolution images independently of weather conditions. Its ability to operate in all environmental conditions makes it useful for a variety of applications. However, synthetic aperture radar imagery has a few typical features that cause distortion and affect human perceptibility of the image. The distortions such as speckle, layover and diffusion degrade the overall image quality and complicate the interpretation of an image. To address these challenges, multiple enhancement methods have been developed, varying from basic filtering to deep learning-based super-resolution models. Evaluating the effectiveness of such methods is vital and for that purpose, several image quality assessment metrics have been developed. However, many of these metrics created are intended for optical imagery and they do not translate well to synthetic aperture radar imagery due to the nature of the imaging methodology.

This thesis focuses on investigating image quality assessment methods and metrics in the context of synthetic aperture radar super-resolution images. The research is done by following the guidelines of iterative literature review, which operates by dividing the study into iteratively executed cycles. The main idea is to include and exclude sources in iterative fashion by proceeding further into each source as rounds go by. This methodology makes it possible to go through many existing research in timely manner. The search is done using keywords, and each subtopic is done separately. The sources that emerge during this process, are further categorised by their respective subtopics.

The findings suggest that many of the existing metrics struggle with synthetic aperture radar-specific disruptions, and that no individual metric can assess synthetic aperture radar super-resolution image quality effectively. This is due to the variety of applications of synthetic aperture radar imagery and their specific needs in the context of image quality. The existing super-resolution image quality metrics are focused on specific features in the image and can effectively estimate the image quality with respect to that. Since super-resolution image quality metrics are focused and can't estimate the overall image quality alone, it is suggested to utilise multi-metric evaluation strategies. The findings are highlighting the need for further research into multi-metric methodologies for synthetic aperture radar super-resolution.

KEYWORDS: SAR, Super-resolution, Image quality assessment, Quality metrics

VAASAN YLIOPISTO
Tekniikan ja innovaatiojohtamisen akateeminen yksikkö
Tekijä: Jami Vihavainen
Tutkielman nimi: SAR image quality metrics with respect to Super-Resolution
Tutkinto: Tekniikan kandidaatti
Oppiaine: Automaatio- ja tietotekniikka
Työn ohjaaja: Jouni Lampinen
Valmistumisvuosi: 2026 **Sivumäärä:** 50

TIIVISTELMÄ:

Synteettisen apertuurin tutka on tutkakuvantamistekniikka, jolla voidaan tuottaa korkealaatuisia kuvia riippumatta säästä. Sen kyky toimia kaikissa sääolosuhteissa mahdollistaa monia hyödyllisiä käyttötarkoituksia. Synteettisen apertuurin tutkakuvissa on kuitenkin muutamia tyypillisiä ominaisuuksia, jotka aiheuttavat häiriöitä ja vaikeuttavat kuvien ymmärrettävyyttä. Häiriöt kuten speckle, layover ja diffusion huonontavat kuvan yleistä laatua ja vaikeuttavat kuvan tulkitsemista. Näiden vaikeuksien takia on kehitetty monia parantamismenetelmiä yksinkertaisista suodattimista aina syväoppimispohjaisiin superresoluutiomalleihin asti. Tällaisten menetelmien tehokkuuden arviointi on oleellista käytettävyyden parantamiseksi, ja sitä varten on kehitetty monia kuvanlaadun arviointimetriikoita. Monet näistä metriikoista on kuitenkin kehitetty optisia kuvia varten, eivätkä ne sovellu käytettäväksi yhtä hyvin synteettisen apertuurin tutkakuvien kanssa kuvantamismenetelmän ominaisuuksien takia.

Tämä tutkielma keskittyy tutkimaan kuvanlaadun arviointimenetelmiä ja -metriikoita synteettisen apertuurin tutkakuvien yhteydessä. Tutkimus on tehty noudattamalla iteratiivisen kirjallisuuskatsauksen menetelmää, joka toimii jakamalla tutkimus iteratiivisesti suoritettaviin kierroksiin. Pääajatus on sisällyttää ja rajata lähteitä pois iteratiivisesti edeten siten, että joka kierroksella perehdytään lähteisiin yhä syvällisemmin. Kyseinen menetelmä mahdollistaa monien olemassa olevien tutkimusten käsittelyn lyhyen ajan kuluessa. Lähteiden haku tehdään avainsanojen avulla ja jokainen alateema käsitellään erikseen. Prosessin aikana esiin nousevat lähteet luokitellaan teemakohtaisiin kategorioihin.

Työn löydökset viittaavat siihen, että synteettisen apertuurin tutkalle tyypilliset häiriöt aiheuttavat haasteita monen nykyisen metriikan hyödyntämiselle. Löydöksissä havaitaan myös, että mikään yksittäinen metriikka ei kykene arvioimaan superresoluutiokuvien laatua tehokkaasti, koska synteettisen apertuurin tutkalla on moninaisia käyttökohteita ja kuvanlaatuun liittyviä vaatimuksia. Nykyiset superresoluutiokuvien laadun mittarit keskittyvät kuvan tiettyihin piirteisiin ja pystyvät tehokkaasti arvioimaan laatua vain näiden osalta. Koska superresoluutiokuvien laadun mittarit ovat rajattuja eivätkä pysty yksin arvioimaan kokonaislaatua, suositellaan käytettäväksi useampaa rinnakkaista laadun metriikkaa. Työn tulokset nostavat esiin tarpeen lisätutkimukselle monimittaristen menetelmien kehittämiseksi synteettisen apertuurin tutkan superresoluutiota varten.

AVAINSANAT: SAR, Superresoluutio, Kuvan laadun arviointi, Laatumetriikat

Table of Contents

1 Introduction

Synthetic Aperture Radar (SAR) is a type of radar imaging method commonly used in satellite-based earth observation applications. SAR methodology utilises electromagnetic waves which are transmitted in sequences to surface area under observation. The backscattered echoes from the target are received as complex matrix data and reconstructed to form an observable image. In the satellite applications, SAR has many different advantages compared to other conventional imaging methods. The main difference between radar-based methods and conventional optical method is the wavelength. Due to wavelength of the electromagnetic waves, SAR has the ability to collect observations unaffected by prevailing conditions such as time of the day and weather. This creates many intriguing opportunities compared with conventional optical methods.

SAR can be a useful tool, for example in environmental, agricultural and climate monitoring. The ability to monitor ice sheets and glaciers or detect changes in forests and fields is necessary to understand environmental processes, track climate change in long term, and support effective resource management. It also has a variety of maritime, military and environmental applications. Although SAR generates many opportunities, it also has some downsides caused by the very nature of radar imagery. Extremely flat surfaces, such as calm water or a runway on an airfield, can cause specular reflection where most of the energy from the radar reflects away. This induces a dark appearance of the spot. On the other hand, urban areas and complex obstacles frequently exhibit double reflections, which happen when the signal bounces multiple times before returning to the radar. Urban areas and especially tall buildings introduce a unique effect called layover, which can be seen as the top of the obstacle seems to be “shifted” away from its real position. This causes SAR image to have layered object features and can cause challenges, especially in feature recognition. Also, some surfaces such as forests and choppy water can cause the radar energy to reflect in many different directions. This is called diffusion, and it results in strong and bright appearance in the image. Speckle appears as arbitrary grainy pattern and is caused by reflection interference within the SAR pixel. The reflections in each SAR pixel have slightly different phases and when added together they interfere randomly. When these pixels are added together to an image, they form a speckle. Speckles are a distinctive feature of SAR image and cannot be removed completely from the image. Strong speckle patterns can obscure subtle changes in the backscatter making small features more difficult to identify.

There have already been numerous methods created to resolve these problems and enhance the overall image quality. However, each of the methods have been created to target a specific problem regarding image quality. For example, in the case of speckle, the most typically used methods are varying filter-based methods and most recently, deep-learning based methods. For generic image quality like resolution and clarity, the most used methods are either relying on more classical image altering methods such as contrast and interpolation or deep learning image super-resolution (SR) models. To compare analytically the performance of these image enhancing methods, there is a need to have accurate and descriptive quality evaluation metrics. For normal optical images, there already exists various numerical metrics. For example, structural similarity index measure (SSIM) and peak signal-to-noise ratio (PSNR). However, due to different nature of SAR and optical imagery, all these metrics are not fully applicable to SAR images.

This thesis will mainly focus on image quality assessment methods and metrics especially for SAR SR imagery. The objective is to research the most effective metrics and best practices to evaluate the quality and features of SAR imagery improved using SR algorithms.

2 Background

To properly understand the problem that this thesis is discussing, it is first necessary to understand the background revolving around the topic.

2.1 Synthetic Aperture Radar

Synthetic Aperture Radar (SAR) is an imaging technique that is known for its ability to see through clouds and darkness (Ager, 2023). This ability provides numerous intriguing capabilities that are not available with optical imaging methods. This characteristic ability is made possible by the nature of SAR as an imaging method. SAR operates similarly that a conventional radar, where electromagnetic waves are transmitted in sequences and the backscattered echoes are collected by the radar antenna (Moreira et al., 2013). A SAR image is reconstruction from backscattered echoes, and it represents a measurement of the scene reflectivity. The echoes received, form a two-dimensional complex matrix where each sample contains two values, the real and the imaginary which can be seen as an amplitude and phase value (Ager, 2013; Moreira et al., 2013). The first dimension is corresponding to the range direction consisting of amplified and base band converted samples of the complex echo signal. As the radar moves, it forms successive range lines along the direction of travel, also known as the azimuth direction, thereby forming the second dimension of the complex data matrix. The results from processed radar backscatter data represent a measure of the scene's reflectivity more closely than conventional optical satellite imagery does (Moreira et al., 2013).

In the early days of radar imaging, so called real aperture radars (RAR) were invented. With this imaging method, it was not possible to create high-resolution images from a long distance (Ager, 2013). This limitation is because of the resolution in azimuth dimension is equivalent to the beam pattern size on the ground. Image 1 demonstrates that RAR is reliant on the beamwidth (β). The beamwidth widens when going from near to far range, and size of ground resolution cell is the beamwidth times the range (Ager, 2013). For this reason, RAR had azimuth resolution, in the hundreds of meters, and it could not produce high-resolution images.

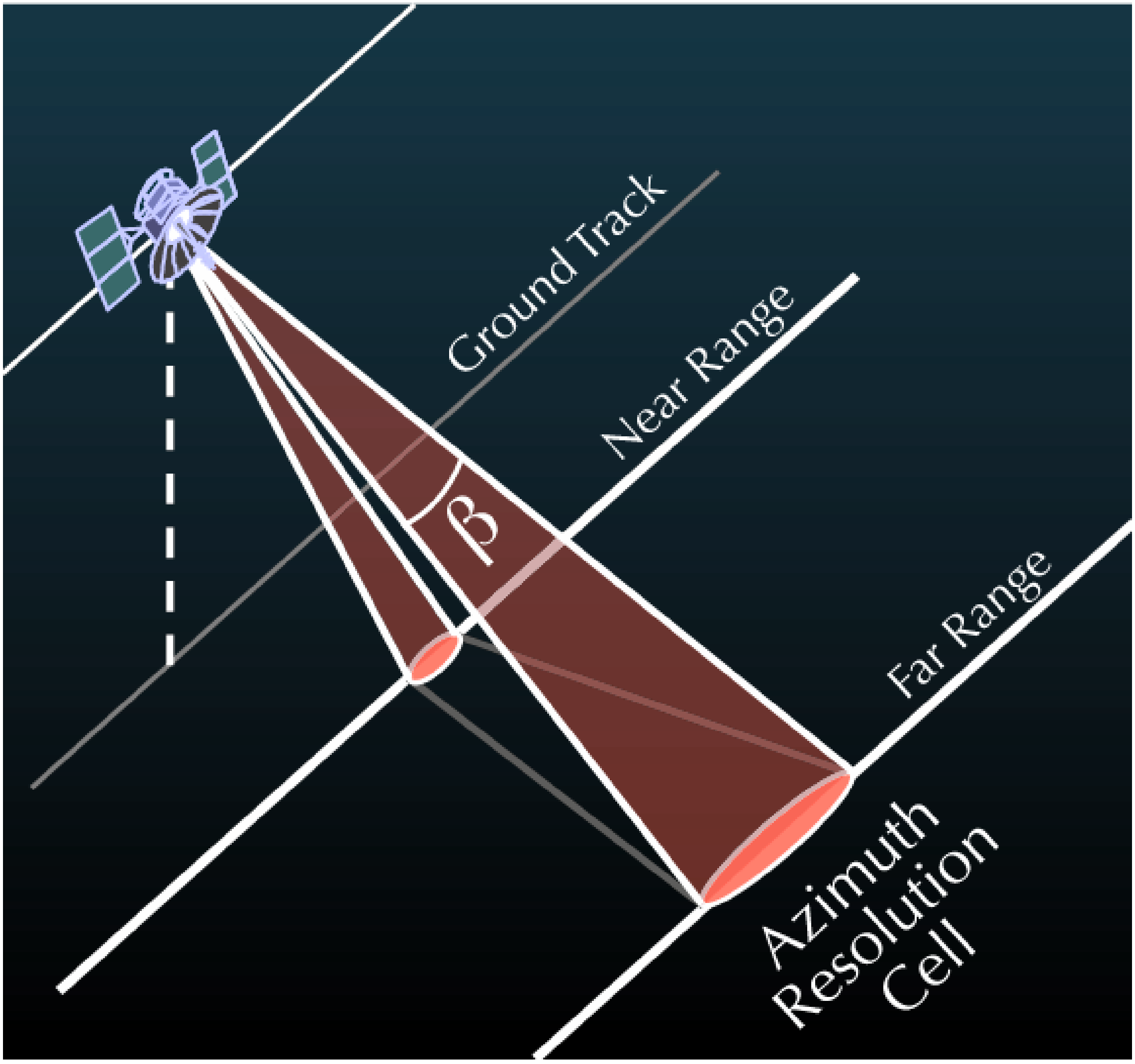

Image 1. Azimuth resolution for real aperture radar (Ager, 2013).

Since the beamwidth depends on the antenna width in the azimuth direction:

$$Azimuth\ beamwidth\ \beta = \frac{\lambda}{d}, \tag{1}$$

where λ is radar wavelength and d is antenna width, it can be determined that increasing the antenna size can improve the resolution of the image (Ager, 2013). However, building antennas that big is not realistic, and so invention of SAR technology was ground-breaking at the time. With SAR methodology, a combination of the backscattered echoes and movement allows the construction of a synthetic aperture that is significantly longer than the physical antenna length. Because of the synthetic aperture, SAR can create high-resolution images independent of the distance to the target (Ager, 2013). The azimuth resolution can be calculated with the wavelength of the microwave energy and the angle subtended by the synthetic aperture:

$$SAR\ azimuth\ resolution\ = \frac{\lambda}{2\Delta\theta}, \tag{2}$$

where λ is radar wavelength, $\Delta\theta$ is angular extent of the synthetic aperture as the radar moves on its path (Ager, 2013). This gives SAR its characteristic ability of being independent of distance. All the measurements completed during data collection, are treated as they are collected simultaneously using one very long antenna (Ager, 2013). This eliminates the need for physically long antenna and enables the collection instruments to be housed by a smaller satellite.

The nature of SAR causes the features within the images to not be geometrically accurate. This is happens due to the radar only measuring the three-dimensional scene in the radar coordinates slant-range, meaning the direct distance from the radar to the target, and azimuth (Moreira et al., 2013). This causes many unusual geometric effects such as foreshortening and layover which can be seen as compressed or stretched parts of the image in places of sloped terrain (Moreira et al., 2013). Foreshortening is a phenomenon where sloped terrain is compressed. This happens when a slope is positioned on SAR images based partly on range and it causes the slopes facing the radar to appear as compressed (Ager, 2023). The layover can be seen as an opposite phenomenon compared to the foreshortening. When the top of a tall object appears closer to the radar than its base, the effect is called layover. It happens when the collections range curve is shallower than the slope. The volume of layover effect increases as the collection angle becomes steeper and the incidence angle gets smaller (Ager, 2023). The direction of the layover is perpendicular to the flight path of the satellite when it is straight and level (Ager, 2023). These geometric features are natural effects of representation of three-dimensional objects in two dimensions. They are not only unique to SAR imaging or even radar illumination, but also appear often in optical imagery (Ager, 2023). The collection geometry in space-based SAR satellite is further illustrated in image 2.

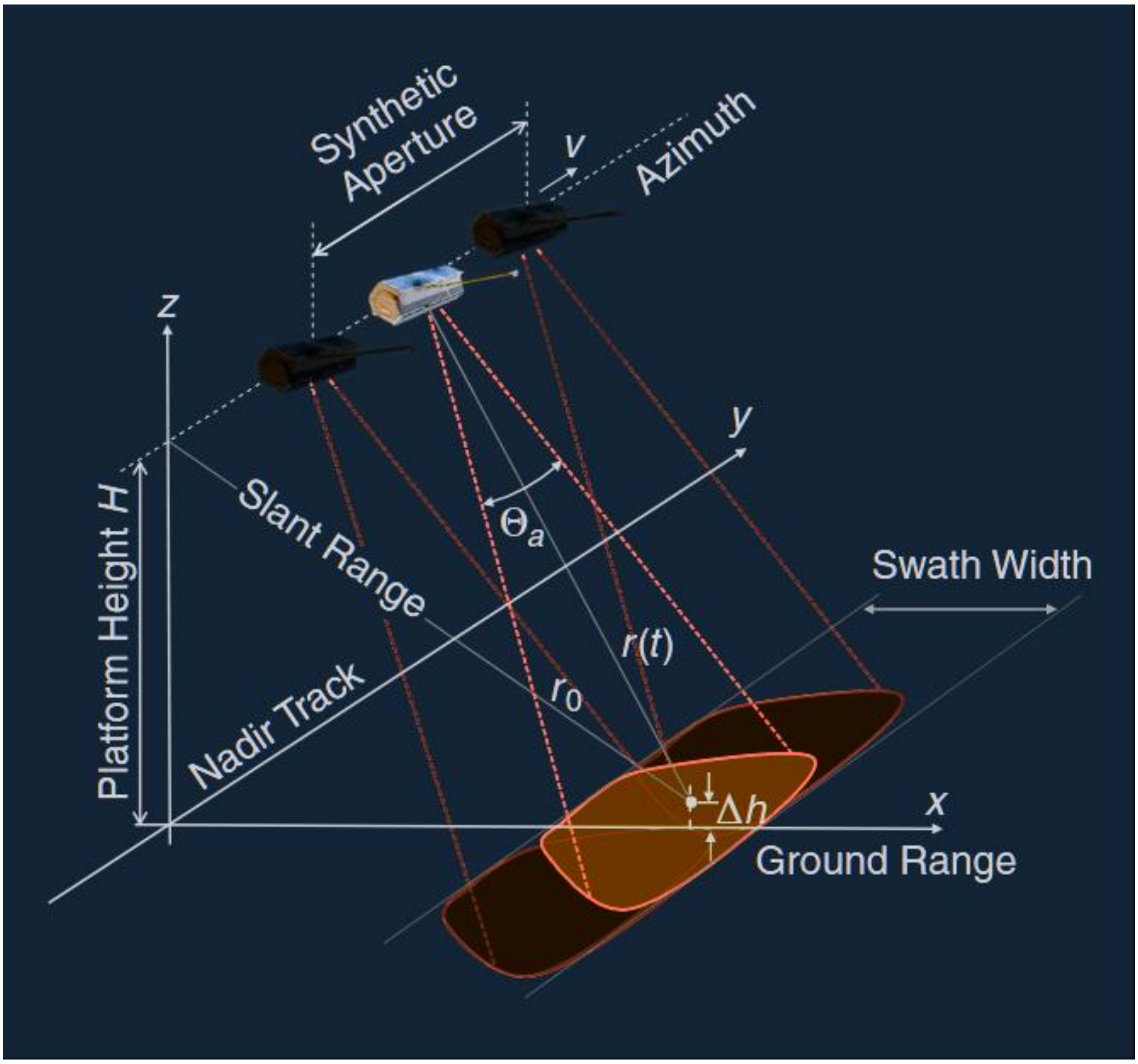

Image 2. Illustration of SAR imaging geometry (Moreira et al., 2013).

Pixel superimposition is an effect typically associated with the layover distortion in SAR imagery. The effect is caused when scatter of equal range is interacting with different objects in different ways (ICEYE, 2024). This can be seen as objects being “transparent” with the ability of seeing different objects behind them. The effect happens when two different backscattering echoes return to the radar the same time but from different targets. The reflections returning at the same time causes their values to be merged in the acquisition process, hence creating a medium pixel of the backscattered echoes that can be interpreted as transparent in the SAR image (ICEYE, 2024). This effect is further illustrated in image 3, where specular diffusion (black pixel) and diffuse reflection (light pixel) are forming their medium pixel (light grey) which creates the appearance of partial transparency (ICEYE, 2024).

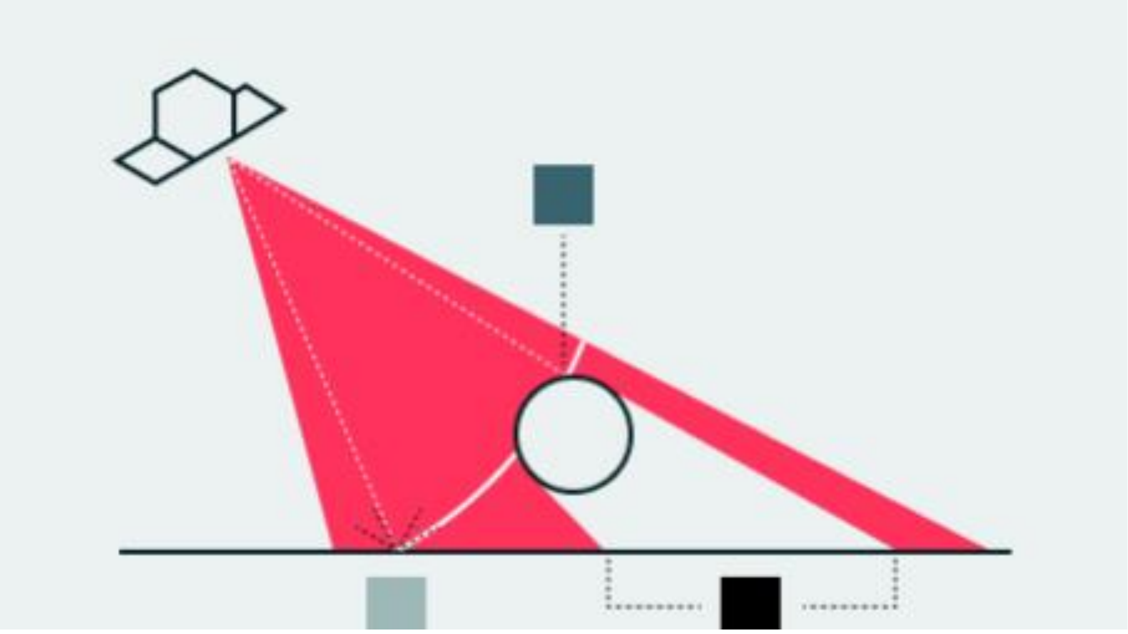

Image 3. Superimposition of pixels (ICEYE, 2024).

While foreshortening and layover are distorting the image, the shadow is formed when an area is directly behind an object, hence not appearing in the image. This is highly due to the incidence angle in which the image itself has been captured. In image 4, the effects of changing the incidence angle θ can be seen affecting the visual presentation of the scene. In images 4.a and 4.b, the change in the incidence angle creates compression or stretching while in image 4.c the same part is invisible to the radar creating the so-called shadow effect. The shadow effect appears in the SAR image as a completely black area as contrast to the dark shadows on optical images (Ager, 2013). These areas can still contain signal noise and other image features even though they are considered null due to lack of backscattered echoes.

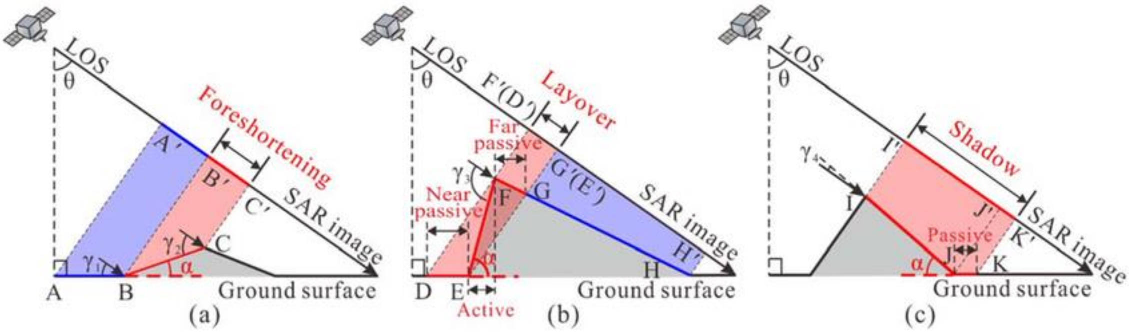

Image 4. SAR imaging geometry features (Ren et al., 2021)

Speckle is a characteristic feature of SAR imagery. It appears as random grainy pattern, and it can be seen in every SAR image. The speckle noise is caused by interference of returning signals from many reflectors located within one resolution element (Tuzova et al., 2020). The speckle is usually considered as multiplicative noise but is an inherent feature in SAR imagery as it can contain significant information about the observed scene (Dellepiane & Angiati, 2014; Tuzova et al., 2020). Speckle is also typically referred as noise. However, it cannot be reduced by increasing signal power because its variance increases with its intensity (Moreira et al., 2013). Speckle typically appears in places within the scene where the radar wavelength is comparable to the surface roughness with distributed scatterers (Moreira et al., 2013).

These image features are essential to understand before interpreting SAR imagery. Each of the disruptions create a different type of problem and are each combatted in different ways. Disregarding their differences, they are all important, characteristic and heavily contributing to human perceptibility of SAR images.

2.2 Image super-resolution

Image super-resolution (SR) is one of more modern application of image quality enhancement. The objective for image super-resolution is to reconstruct a high-resolution (HR) image from a single low-resolution (LR) image or a sequence of images (Nasrollahi & Moeslund, 2014; W. Yang et al., 2019; Yue et al., 2016). SR methods can be roughly split into two categories: frequency domain and spatial domain. Historically, the first super-resolution methodologies emerged from signal processing techniques in the frequency domain (Nasrollahi & Moeslund, 2014; Yue et al., 2016). However, most of the recent SR methodologies operate mainly in the spatial domain.

SR techniques can be further divided into groups called single-image super-resolution (SISR) and multi-image super-resolution (MISR). This division is concluded based on the number of images utilised in the construction of single HR image (Nasrollahi & Moeslund, 2014; W. Yang et al., 2019). Both categories create their own challenges and opportunities. If multiple images with minor pixel misalignment can be acquired from the same site, the information can be used to create HR image of the scene (Yue et al., 2016). The MISR algorithms typically assume relation in geometric or photometric displacements from the targeted HR image. These algorithms usually use to their advantage the differences between LR observations to reconstruct the targeted HR image, hence they are also often referred as reconstruction-based SR techniques (Nasrollahi & Moeslund, 2014). However, acquiring multiple images that are only minorly different can be difficult and that is why many of the SR methodologies are revolved around single-image techniques. SISR methodologies are also highly efficient hence, they are generally more popular than MISR methodologies (W. Yang et al., 2019). The limitations of SISR techniques come from the limited amount of HR reconstruction data per LR image. Due to the lack of HR data, most of the SISR based algorithms are basically forced to hallucinate and invent the missing information of the super-resolved images utilising the HR images from a training image collection (Nasrollahi & Moeslund, 2014).

2.2.1 Frequency domain

Super-resolution algorithms in the frequency domain first transform the low-resolution image to the frequency domain where the high-resolution image is further estimated (Nasrollahi & Moeslund, 2014). After the estimation, the high-resolution image is transformed back into the spatial domain. Depending on the chosen transformation for the image frequency domain conversion, these algorithms can generally be divided into Fourier-based and Wavelet-based transform methods (Nasrollahi & Moeslund, 2014).

The main premise of Fourier-based transformation is to exploit multiple LR images of same scene as their differences contain information due to subpixel displacements. The first Fourier-based SR model expected every LR image to be a shifted version of the observed scene and their relationship is modelled as:

$$g_k(m,n) = f(mT_m + \Delta m_k, nT_n + \Delta n_k), \qquad (3)$$

where T_m and T_n are the sampling periods of the LR image, and Δm_k and Δn_k are relative shift properties (Nasrollahi & Moeslund, 2014). The transformed LR images can be further related to the HR image due to the Fourier shift property. This relationship can also be written as:

$$G = \Phi F_f, \qquad (4)$$

where Φ relates the discrete Fourier transform of the LR images G to the continuous Fourier transform of the HR scene F_f (Nasrollahi & Moeslund, 2014). This assumed ideal LR images with no blurring effects, which is effective providing the basis of recovering the HR image through spectral estimation. However, real-life images are affected by noise, blur and other inaccuracies. This is why later an additive noise, and blurring effects were added to the method, and it was rearranged as:

$$G = \Phi F_f + \eta, \qquad (5)$$

where η is the noise term (Nasrollahi & Moeslund, 2014). Further research addresses unpredictability in the subpixel displacements. As the displacement parameters are not known precisely, the system model was modified. To incorporate the errors of estimating the displacements between the LR images a recursive total least squares method was included, hence the equation becomes:

$$G = (\Phi + \mathrm{P})F_f + \eta, \qquad (6)$$

where P is a perturbation matrix constructed from the estimation errors (Nasrollahi & Moeslund, 2014).

Wavelet transform is generally considered as an alternative to the Fourier transform and it is widely used in frequency domain-based SR algorithms (Nasrollahi & Moeslund, 2014). Typically, it is utilised to split the input LR image into correlating sub-images which allows exploiting the similarities between neighbouring regions. SR methods based on Wavelet may have inefficiencies in implementing degraded convolution filters, while they can be done efficiently with the Fourier transform. This is why these two transformations have sometimes been combined into the Fourier-wavelet regularised deconvolution.

2.2.2 Multi-Image Super-Resolution

Super-resolution models that are categorised in the multi-image category have a common approach to the training data. So called multi-image super-resolution (MISR) is a category of super-resolution models that utilise two or more images in the training process. Also, most of the MISR algorithms use estimated motion to relate the LR input frames (L. Pickup, 2007). Based on these criteria, the methods in frequency domain such as Fourier and Wavelet transformations can be also counted into the multi-image category. These methods work similarly to other MISR algorithms as their approach is to model shifts between LR images (Afrasiabi et al., 2023). However, they work in the frequency domain, hence they are typically put into their own separate category. Other MISR algorithms usually work in the spatial domain.

Iterative back projection methods are MISR reconstruction algorithms that utilise reverse-engineering to construct HR image from LR images. They set and initial estimate of the HR image and then try to refine that by using iterative optimisation process (Afrasiabi et al., 2023; Irani & Peleg, 2003). (Irani & Peleg, 2003) present an iterative reconstruction algorithm which begins from the initial estimate of HR image and continues with simulating imaging process to construct corresponding LR images. Then difference is calculated between these simulated images and the initial LR inputs which results in difference images. The difference images are then used to improve the initial HR estimate by "back-projecting" the reconstruction error to the HR image. This process is iteratively repeated to minimise the reconstruction error:

$$e^{(n)} = \sqrt{\sum_k \sum_{(x,y)} (g_k(x,y) - {g_k}^{(n)}(x,y))^2}, \tag{7}$$

where $e^{(n)}$ is the total error, $g_k(x,y)$ is the value of the pixel at location (x,y), and ${g_k}^{(n)}(x,y)$ is the simulated pixel value from the current HR estimate (Irani & Peleg, 2003). Another example of reconstruction-based SR framework is the Baylean Maximum a Posteriori (MAP) approach. The Baylean equation is typically solved using the MAP framework which combines data fidelity with a prior on the HR image (L. C. Pickup et al., 2008). Bayesian based SR algorithms define SR as a probabilistic inverse problem. To estimate the HR image, prior assumptions considering image structure such as smoothness and edge preservation are used along the observed data. Because maximum-likelihood formulation is highly sensitive to noise, the MAP formulation is the dominant reconstruction framework in SR (L. C. Pickup et al., 2008). Back-projection algorithms can be seen as an error correction solution, whereas Bayesian MAP based reconstruction algorithms utilise data fidelity and prior knowledge, and are able to generate more general estimation.

Deep learning methods are not as developed in multi-image solutions as they are in single-image solutions, however some notable architectures have been developed (Afrasiabi et al., 2023). The method known as DeepSum, uses convolutional neural-network (CNN) based architecture and utilises spatial and temporal correlations between LR images (Afrasiabi et al., 2023; Bordone Molini et al., 2020). The framework integrates the spatial registration task inside the CNN, which allows enhanced registration accuracy (Bordone Molini et al., 2020). Their SR process relies on single CNN with three main stages that each have their own purposes. The SR model described is following a supervised deep learning approach where the CNN is utilised in learning the residual between bicubic interpolation and the ground truth. A preprocessing step is done before entering the first stage. Preprocessing executes a bicubic interpolation to desired size for the LR images before feeding the results in the CNN. The first phase acts as feature extractor, responsible for extracting high-dimensional features from the interpolated LR set of input images. In this stage each of the input images is processed separately by a sequence of 2D convolutional layers. Its goal is to exploit spatial correlations to improve the initial bicubic interpolation. The second phase aims to estimate a set of filters derived from the previous high-dimensional feature representations. The phase utilises the features to compute optimal registration for each image instead of performing it in image space. The third phase merges the image representations produced by the second phase, by exploiting a sequence of 3D convolutional operations with small kernels (Bordone Molini et al., 2020). Following this approach allows the registration task to utilise learning capabilities of the network to increase accuracy and create resiliency to scene variations.

2.2.3 Single-Image Super-Resolution

Single-Image Super-Resolution (SISR) is a category of super-resolution models that can include every SR model that uses a single LR image as the input. Generally, solving SISR algorithm is an ill-posed problem because single LR input may correspond to multiple HR solutions (Wang et al., 2021; W. Yang et al., 2019). Typical SISR methodologies can be modelled as:

$$y = (x \otimes k) \downarrow_s + n, \tag{8}$$

where $x \otimes k$ is the convolution between the blurry kernel k and the HR image x, $\downarrow_s$ is the down sampling operator with scale factor s, and n is the independent noise term (W. Yang et al., 2019). Typical SISR algorithms are essentially split into three categories: interpolation-based methods, reconstruction-based methods and learning-based methods (W. Yang et al., 2019).

Image interpolation refers to resizing digital images and is used in many applications related to images (Wang et al., 2021). Many different interpolation methods exist, and they are still widely used in CNN-based SR models due to their interpretability and ease of implementation. Nearest-neighbour interpolation selects the value of the nearest pixel for each position to be interpolated disregarding other pixels (Wang et al., 2021). This method is extremely fast but typically the results are blocky and low quality. Bilinear interpolation performs linear interpolation in two axis separately. It results in quadratic interpolation with a receptive field sized 2 x 2, which produces better quality results than nearest-neighbour interpolation while maintaining relatively fast speed (Wang et al., 2021). Bicubic interpolation performs interpolation on the two axis similarly than the bilinear interpolation. Compared to bilinear interpolation, the bicubic method takes into account 4 x 4 area of pixels, resulting in smoother image with fewer disruptions but is considerably slower (Wang et al., 2021). Bicubic interpolation is the mainstream method for pre-sampling and building SR datasets, where it is typically used to degrade HR images to LR for model training purposes. Generally, interpolation-based upsampling methods use the images own image signals without considering any outside-information to improve the image resolution.

An example of reconstruction based SISR model is the sparse-representation (J. Yang et al., 2010). In this method two constrains are considered: HR image produced by the model should be consistent with the LR input, and sparsity prior, which assumes that the HR patches are sparsely representable in chosen dictionary and can be recovered from the LR input. The main presumption is that a HR patch and a comparable LR patch can be represented by the same sparse coefficient vector (J. Yang et al., 2010). The algorithm itself works by taking each LR patch and proceeds to compute a sparse code with respect to the LR dictionary. After that it applies the same coefficients to the HR dictionary, predicting a sparse representation from the LR patch and then transferring it to HR space. This reconstruction process has two stages. First stage is where a local patch model reconstructs HR patches separately from sparse coding. The LR input is presented using features highlighting high-frequency content. Second stage is where the prior HR image is refined with a global reconstruction constraint, which projects results back to an image that has down sampled version matching the LR input. In this design the sparsity covers important details, and the reconstruction constraint maintains consistency with the observation model (J. Yang et al., 2010). This model can be included both in classical reconstruction and more modern learned SR. While the LR/HR relationship is learned from the training data, the model can still be classified as reconstruction-based due to the model's explicit constraint on LR image reproducibility.

SRCNN is a good example of one of the earlier CNN-based SISR models (Dong et al., 2015; Wang et al., 2021). SRCNN performs a mapping beginning from a LR input to a HR image, using a convolutional network to replace the stage in older methods. It is argued that traditional sparse-representation frameworks can be modelled utilising CNN with the key difference of stages being processed together rather than individually (Dong et al., 2015; Wang et al., 2021). The model itself follows a simple concept. Firstly, the LR input image is processed using bicubic interpolation, after which the CNN refines these scaled images. Within the CNN, the first layer is used to extract features from the image, the middle layer uses the extracted features to map them into HR feature space, and the final layer generates the new HR image from these transformed features (Dong et al., 2015). This model demonstrates that a simple CNN can generate high quality results that are outperforming many of the earlier interpolation and sparse-coding based models in terms of quality. Other CNN-based models developed later have improved the SRCNN by various methods. For example, transferring most of the computation to low-resolution space, widening the network or adding residual learning. Still SRCNN can be seen as the foundation for CNN-based super-resolution models (Wang et al., 2021).

SRGAN is another example of SISR model utilising deep learning (Ledig et al., 2017). Whereas the objective for SRCNN is to achieve perfect pixel-wise reconstruction of LR image, the SRGAN is aiming to optimise super-resolution for perceptual realism. Ledig et al. (2017) argue that models trained mainly using the mean squared error tend to have higher PSNR, but they generate overly smooth images with weak high-frequency details. The design of a GAN consists of a generator generating the images, and a discriminator that tries to distinguish generated images from HR images (Ledig et al., 2017; Wang et al., 2021). During the training of GAN model, these two are performed alternately resulting in generator producing consistent outputs that the discriminator can’t distinguish from real HR images (Ledig et al., 2017; Wang et al., 2021). The generator itself is built as a deep residual network, which is trained using perceptual loss that is built with a VGG feature map-based content loss and an adversarial loss function. When comparing SRGAN with SRCNN, SRGAN stands out with its target of improving visual realism at the cost of PSNR, whereas SRCNN is mainly a CNN regression model focused on pixelwise accurate reconstruction and maximising the PSNR (Wang et al., 2021). SRGAN can be seen as the foundation for GAN-based single-image super-resolution models.

3 Research methodology

To gain knowledge about existing quality evaluation methods applicable to super-resolution processed SAR imagery, it is first necessary to research the topic. The choice of methodology for this thesis is based on existing research on the topic. As this topic has not been deeply investigated in the context of SAR imagery with respect to super resolution, conducting a comprehensive literature review is necessary.

The aim of this review is to establish the current state of research and to identify potential gaps that could guide future studies. This work therefore provides a foundation for further research on the topic.

3.1 Research plan

The research process follows guidelines of iterative literature research that are documented in table 1. The method divides the research process into multiple rounds. In the first round, sources are selected and categorised based on their title and abstract.

The second round involves a deeper examination of these sources by reviewing their introduction and conclusion. Sources deemed irrelevant to the topic are excluded from further consideration. In the final round, the remaining sources are read in full, allowing for a comprehensive evaluation of their relevance and applicability. By following this iterative approach, the most relevant sources can be systematically identified over time. This process enables processing a lot of material with great efficiency and quality. In this thesis the sources are also categorised according to the subtopics around the research area.

Iteration	Process	Description
1	Initial screening	Gather sources by topics and categorise them by analysing header and abstract.
2	Further research	Analyse sources further by reading introduction and conclusions. Exclude irrelevant sources from the final screening.
3	Final screening	Read and analyse the source in full and decide whether it is to be included in the final work.

Table 1. Steps of the research process.

The search for sources is conducted by using different keywords and terms to find out answers to pre-determined research questions. This research focuses on identifying the factors that influence SAR image quality and further determining the most suitable quality assessment metrics for SAR super-resolution images. These focuses also serve as the main research questions of the thesis. The aim of the research process is to identify and gather sources that best address these questions. The inclusion and exclusion of sources on each iterative round is based on how well each source contributes to answering to the questions. On each round the sources are evaluated based on their relevance to the topic, retaining only the most relevant and informative sources to be further analysed. This iterative evaluation process is repeated for each subsection related to the topic of this thesis with respect to the research questions.

3.2 Research process

Analysed sources were gathered using several repositories. Most of the sources were found through either Google Scholar or from the library of University of Vaasa. Many of the sources are articles, research papers and scientific publications but also some books and websites are included. During research process, the topic was divided into few different categories, and each category was examined separately. The defined main categories were SAR image quality, super-resolution image quality and SAR super-resolution image quality. Besides these categories, separate searches were done for subtopics like SAR and super-resolution. The main categories were focused strictly on the topic and the research question, while the additional categories were focused on gathering supporting information. The search was done by utilising keywords and different combinations and variations of them. For the main categories, the most effective keywords for surfacing sources relevant to the topic were “SAR image quality metrics” and “SAR super-resolution image quality”. The search parameters used were re-visited and refined during the research process based on the existing sources and the insights they provided. Based on the results gained using these parameters, the sources were selected for inclusion in the iterative process. As described in chapter 3.1, the first iteration focused on the title and abstract of the source. The purpose of this round was to conduct an initial evaluation of each source’s relevance and based on this assessment, either include or exclude it from further analysis. This enabled the efficient exclusion of a large number of sources based on their general relevance to the topic. In addition to assessing relevance, this round enabled a critical evaluation of each source’s credibility. The second round focused on a more detailed assessment by reviewing the introductions and conclusions of the selected sources. This round made it possible to exclude sources that initially seemed relevant but after closer examination can be considered to be too far removed from the original topic. The sources retained after the second round were then examined in full during the third round. This final screening stage ensured that the sources were relevant throughout the entire paper. Only the sources passed from the final stage were included into the literature basis of this thesis.

4 Super-Resolution image quality metrics

To effectively compare image quality, it is necessary to decide on what metric to use. The decision between different metrics is heavily influenced by the use-case and available resources. As SR models have been developed rapidly during the recent years, there has also been a variety of quality metrics presented. These quality metrics are typically categorised based on their goal or reference availability. In this thesis they will be categorised based on reference availability, creating 4 different categories. First category is the subjective category, where the quality of image is based on human opinion. Second category is the full-reference (FR) category, which includes metrics that are comparing SR output with a ground-truth HR image. Third category is the reduced-reference (RR) category, where models utilise reference information partially (Mohammadi et al., 2014). The last category is the no-reference (NR) category, which consists of models that are analysing image quality without using the ground truth. The metrics placed in FR, RR and NR can all be also categorised as objective image quality metrics. The objective for this chapter is to walk through these different categories of these image quality metrics and demonstrate both common and important metrics in their own fields.

4.1 Subjective metrics

As the objective for super-resolution can be defined as generating high-quality images from low-resolution images, human perception is an essential part of the process. The most accurate and reliable way of assessing image quality is subjective evaluation, since humans are the ones who are working with the end-results (Mohammadi et al., 2014). This is why human opinion must be included in the quality assessment process. Mean Opinion Score (MOS) is a simple metric that reflects the human opinion effectively. MOS is widely used in human subjective studies, and it is obtained by manually scoring each image by different people (International Telecommunication Union, 2013; Ma et al., 2021). After that a mean is calculated to determine the overall opinion on each image.

This method focuses heavily on human perception, but it is slow and challenging to accomplish due to the amount of people needed to give their scores.

Difference Mean Opinion Score (DMOS) is another subjective quality metric, which is based on a reference rather than a standalone quality rating. Similar to the MOS, DMOS uses human ratings to construct the overall rating each image. Instead of applying the rating directly, it uses the difference between images and is defined as the difference between the scores of reference and test images (Mohammadi et al., 2014). DMOS helps to reduce variance, and can achieve better quality estimate than raw MOS.

Pair-wise comparison is more a type of subjective image quality method rather than a metric. Pair-wise comparison works by showing observers two images of the same scene, from which they are asked to choose the higher quality one (Mohammadi et al., 2014). Observers are allowed infinite time to make their decision, and they are always required to choose one image from the pair. The challenge is that this method takes more trials to compare each pair. This procedure can also be done with the observers indicating the amount of variety between the images in addition to the initial pick of the better image (Mohammadi et al., 2014). This method provides more useful information of the images for them to be ranked more precisely.

4.2 Full-reference metrics

Full-reference metrics is a category of image quality metrics that consist of metrics that utilise reference material, where the perfect reference image is fully available for quality prediction (Mohammadi et al., 2014). This category varies from mathematical models and pixel difference comparison, all the way to models that use deep learning to assess the quality of image.

Mean squared error (MSE) is a widely used full-reference image quality metric. It reflects the average squared pixel difference between the reference and test images, and it can be calculated as:

$$MSE = \frac{1}{WH} \sum_{j=1}^{H} \sum_{i=1}^{W} (I_{ref}(i,j) - I_{tst}(i,j))^2, \tag{9}$$

where $I_{ref}(i,j)$ is the reference image pixel, $I_{tst}(i,j)$ is the test image pixel, and $W\ and\ H$ represent the width and height of images (Mohammadi et al., 2014). The metric models the pixel-wise error and represents its distortion within the image. Peak-signal-to-noise ratio (PSNR) can be calculated using the MSE, and it represents the power of a signal and distortion. It can be calculated as:

$$PSNR = 10 \log(\frac{D^2}{MSE}), \tag{10}$$

where D is the dynamic range of pixel intensities (Mohammadi et al., 2014). Both MSE and PSNR are simple and computationally cheap methods, hence they are widely used in SR applications. They are effective metrics when considering pixel-wise differences, but tend to perform poorly when predicting human perception of image quality (Mohammadi et al., 2014). This is caused because important characteristics of human visual system are not considered by these metrics.

The structural similarity index (SSIM) is an algorithm that represents the structural information of a scene. This algorithm tries to model the structural information of a scene, attempting to mimic the human visual system which is adapted for extracting this information (Mohammadi et al., 2014). The algorithm itself is based on pixels demonstrating strong dependencies which are carrying useful structure-based information of the scene. Therefore, method utilising this information can create a valid approximation of perceived image distortion (Mohammadi et al., 2014). The structure of objects in a scene is independent of luminance and contrast, which is why SSIM separates the effect of illumination for the structure to be extracted. SSIM algorithm is constructed from three different similarity measurements. First, is comparison of the luminance difference between images. Second, a contrast comparison function that uses standard deviation. Lastly, the structure of each image signal is compared, after which an overall similarity measure is calculated by combining these three functions (Mohammadi et al., 2014). The structural similarity index is defined as:

$$SSIM(I_{ref}, I_{tst}) = [l(I_{ref}, I_{tst})]^{\alpha}[c(I_{ref}, I_{tst})]^{\beta}[s(I_{ref}, I_{tst})]^{\gamma}, \tag{11}$$

where the $\alpha, \beta, and\ \gamma$ are chosen positive constants representing relative importance of each component, $l(I_{ref}, I_{tst})$ is luminance comparison function, $c(I_{ref}, I_{tst})$ is contrast comparison function, and $s(I_{ref}, I_{tst})$ is structure comparison function (Mohammadi et al., 2014). SSIM can be seen as one of the base quality metrics, and it is widely used in many SR related papers as a trusted metric. Based on this metric, there also exists a mean SSIM (MSSIM) index that evaluates the overall quality. This algorithm works similarly to SSIM but at the end, it calculates the mean which produces a single value that evaluates the overall quality (Mohammadi et al., 2014).

SSIM algorithm is considered a single-scale method which achieves the best performance when used at an appropriate scale (Mohammadi et al., 2014). The right scale depends heavily on the viewing conditions, which the algorithm is unable to adapt to. This factor is why a multi-scale structural similarity index (MS-SSIM) has been developed. These kinds of multi-scale methods have an advantage over single-scale methods due to their ability to incorporate image details at different resolutions, and viewing conditions into the quality assessment algorithm (Mohammadi et al., 2014). The algorithm takes in the reference and test images, and performs low-pass filtering and downsampling in an iterative manner (Mohammadi et al., 2014). The same contrast and structure comparison functions used in SSIM are calculated at each scale, whereas the structure comparison function is only calculated at the final scale.

Feature similarity index (FSIM) uses the fact that the human visual system understands an image mainly because of its low-level features like edges and zero crossings (Mohammadi et al., 2014). For the algorithm to assess image quality, it uses two different features. Points with high phase congruency (PC) can be highly informative to the human visual system hence, it forms the main feature of the FSIM algorithm. The PC is also used for weighting, so that structurally important features contribute more towards the final value. PC is contrast invariant and human perception is also affected local contrast of the image, which is why the image gradient magnitude (GM) is the complementary feature of the algorithm (Mohammadi et al., 2014). The FSIM itself is calculated in two stages formed by computation of the PC, the GM, and the similarity measure between the reference and test images. The PC calculations are done utilising the fact that Fourier components with maximum phase include the details that are recognised by the human visual system. The GM is calculated using three different gradient operators that are responsible for obtaining the local contrast information (Mohammadi et al., 2014). Finally, the similarity measure is computed between the PC and the reference, and the GM and the reference, forming the final feature similarity measure index value.

Visual information fidelity (VIF) is modelling natural images in the wavelet domain using gaussian scale mixtures (Mohammadi et al., 2014). If image representing a natural environment has been taken with high quality device that operates in visual spectrum, it is classified as natural scene. VIF algorithm itself consists of three different phases. Firstly, the source model which is designed to model each sub band of the wavelet decomposition as a gaussian scale mixture random field (Mohammadi et al., 2014). The coefficients of each sub band are grouped into blocks that are not overlapping each other. Gaussian scale mixtures that are used by this algorithm can be defined as a random field which is a product of two separate random fields. Secondly, the distortion model is designed as signal attenuation and additive noise in the wavelet domain. Lastly, the human visual system model, is designed as a distortion channel, adding noise to the input, which limits the amount of information flowing through the channel. The visual noise is characterised as a zero mean additive white Gaussian noise modelled in the wavelet domain (Mohammadi et al., 2014).

Most apparent distortion (MAD) is an algorithm that assumes human visual system to use different methods to judge the quality of an image (Mohammadi et al., 2014). The MAD quality metric uses the main idea of humans not looking all distorted images in the same way. If distortions are small, observers typically look at visible errors, while large distortions cause observers to concentrate more on the image's subject matter (Mohammadi et al., 2014). The MAD framework uses this to their advantage by combining a detection-based model for images of better quality with an appearance-based model for lower-quality images. The detection-based model first determines the locations of visible distortions, after which the relationship of pixel values and physical luminance of display is transformed to luminance representation. Then a contrast sensitivity function and a local masking model based on luminance and contrast are applied. Then the representation is combined with a local MSE map in the lightness domain, which highlights error in places where it is expected to be seen (Mohammadi et al., 2014). The appearance-based model is measuring the amount of distortion change in the appearance of the image content. To achieve this, the algorithm utilises log-Gabor filter responses, after which a local statistical difference map is calculated. After both models are computed, these two values are combined to get an measure of perceived distortion (Mohammadi et al., 2014). The final MAD score is a weighted geometric mean e.g. the relation between these two models.

Learned perceptual image patch similarity (LPIPS) is a full-reference perceptual quality metric that compares a test image with a reference image by measuring distance in the feature space of a deep neural network (Karnatov, 2025; Zhang et al., 2018). The primary idea of LPIPS is based on the fact that two images can be considered as similar if their deep feature activations are similar. These feature activation maps are extracted using pre-trained CNN networks that are usually trained for high-level tasks or specifically trained on datasets containing human perceptual similarities of images (Karnatov, 2025).

LPIPS is computed by first extracting feature maps from the CNN which are normalised across channels, and the distance of the reference and distorted image features is computed. Then the resulting distances are weighted and averaged across spatial dimensions to produce a single LPIPS value (Karnatov, 2025; Zhang et al., 2018). This metric has shown to have a high correlation with human perception, and with the deep features it can capture complicated visual structures (Karnatov, 2025). This metric is highly dependent on the dataset used for evaluation and calibration, and the performance can vary due to it.

Deep image structure and texture similarity (DISTS) is a full-reference quality metric that compares distorted image with a reference image in deep feature space (Ding et al., 2020). It operates by combining structure similarity and texture similarity with deep learning context, and it is tolerant to texture resampling. DISTS uses CNN to transform reference and test images to a perceptual representation. With the representation, it estimates structure similarity using correlations of the feature maps. Then it estimates textural similarity with correlations of the spatial averages from the feature maps. The end-result is a score that is designed to preserve meaningful structural changes simultaneously being tolerant to texture variations. A main concept is that the spatial averages of the feature maps are capturing texture appearance, and it can identify two images being perceptually similar even if they are not identical pixel-wise (Ding et al., 2020). This metric is useful especially for restoration tasks, due to it being designed not to judge texture differences in images too harshly if the texture remains perceptually realistic.

4.3 Reduced-reference metrics

Reduced-reference (RR) quality metrics is a category of quality assessment metrics that utilise reference imagery partially. This means that it is not necessary to have access to the original image content for quality evaluation purposes. These metrics only need characteristic information about pixels, transformation coefficients, or other strong features provided (Dost et al., 2022). RR is a practical solution for real-time applications due to reliability and a lack of need for reference imagery, and the algorithms use either relative entropy or entropic difference to their advantage (Dost et al., 2022).

Reduced reference entropic differencing (RRED) is one of the most crucial classical algorithms in the RR quality assessment. The method as an information-theoretic way to assess image quality in the case where the full HR image is missing but there is a small amount of side information available. Rather than comparing pixels directly, it looks at how the information content changes after distortion. The method has predictive performance with adjustable trade-off between accuracy and reference-side bandwidth (Dost et al., 2022; Soundararajan & Bovik, 2012). The algorithm utilises wavelet coefficients obtained using a steerable pyramid decomposition from the reference and distorted image into subbands at multiple orientations and scales (Soundararajan & Bovik, 2012). The Wavelet coefficients of images are modelled using Gaussian scale mixture distributions, separating image content into subbands which allows the method to analyse edges, textures and structural patterns. Then it computes local entropy-based quantities after going through a neural-noise channel. Then it calculates the entropic differences between reference and distorted images and measures the amount of change in the content caused by distortion (Soundararajan & Bovik, 2012). This way the algorithm produces the final RRED quality value.

Reduced-reference structural similarity index (RR-SSIM) is a RR method that estimates image quality through SSIM but with reduced reference (Rehman & Zhou, 2012). The idea is to perform similar SSIM that is done in FR, but with limited reference data. The algorithm begins by feature extraction of the reference image with multiscale multi-orientation divisive normalisation transform (DNT). The algorithm extracts similar features from the distorted image which are then compared to the reference features. Then it estimates a subband distortion with Kullback-Leibler divergence in the DNT domain. Then a structural penalty, based on changes in standard deviations between the reference and distorted images is added. Finally, it estimates SSIM by learning the slope of the relationship between FR SSIM and the RR distortion score using regression-bydiscretization with random forests, which have nearly linear relationship (Rehman & Zhou, 2012). The primary idea of this algorithm is to measure the difference between the reference and distorted images and then transform that RR distortion into an estimate of SSIM.

Spatial efficient entropic differencing for quality assessment (SpEED-QA) is a RR quality metric that is similar to RRED but operates in the spatial-domain (Gupta, 2017). It computes local entropic differences directly from image frames using spatial operations instead of performing heavier wavelet transforms. The algorithm begins with local mean subtraction to the luminance image, creating spatial responses. These responses are then used to produce small local blocks using gaussian scale mixture and a divisive normalisation. After that a neural-noise channel is added and local conditional entropies are computed for corresponding blocks in the distorted and reference images. Then distortion is measured as the absolute difference between these local entropies which is then averaged to produce the quality score (Gupta, 2017). By moving these RRED's core idea to the spatial domain, the algorithm achieves efficiency with only a little loss in predictive power.

4.4 No-reference metrics

No-reference (NR) image quality metrics is a category of image quality metrics that do not require any reference imagery for quality assessment. NR quality assessment is challenging branch of quality metrics due to lack of reference images (Jiang et al., 2022). Current NR metrics are typically separated into two steps. First, they extract distinctive features from the distorted image, and then they use different types of regression operations to map the extracted features to subjective scores (Jiang et al., 2022). The biggest difference between these metrics typically is to do with the extracted features and how they are processed.

Natural image quality evaluator (NIQE) is a NR image quality metric that evaluates the image quality blind. It only uses measurable deviations of statistical regularities in observed natural images without the need of training on human-rated images, and it doesn't need samples of distorted images to construct the model (Mittal et al., 2013). The model measures the amount an image differentiates from the statistical regularities of undistorted images. It is based on constructing collection of statistical features using a simple natural scene statistic (NSS) model. The NIQE starts with preprocessing procedure of the image with local mean removal and divisive normalisation (Mittal et al., 2013). After that, the normalised coefficients are broken into patches from where features are extracted. The normalised coefficients are modelled with generalised gaussian distribution, and the neighbouring coefficients with an asymmetric variant. The computed feature vectors are fitted with multivariate gaussian model to produce a statistical model of test image features. Then this model is compared with a multivariate gaussian model learned from high-quality images, and the distance between them forms the NIQE score (Mittal et al., 2013). This method produces a score that reflects how well local statistics in certain image correspond to those in the high-quality image.

Blind/referenceless image spatial quality evaluator (BRISQUE) is a statistic-based distortion-generic NR image quality assessment metric that works in the spatial domain (Mittal et al., 2012). BRISQUE uses scene statistics of local normalised coefficients to compute the losses of naturalness in the image caused by distortions. The core principle of BRISQUE is that natural high-quality images are obeying natural scene statistic and distorted images are not. The algorithm begins with local normalisation as preprocessing stage. After that NSS features are extracted from the normalised images and further utilised as input to support vector regressor. The support vector regressor is trained on images with associated human quality scores (Mittal et al., 2012). BRISQUE is a classical NR image quality metric that can produce accurate results while staying relatively simple and efficient. The algorithm itself has many similarities to the NIQE metric introduced earlier. The main difference between these algorithms is the material used for model training. Unlike NIQE, BRISQUE is trained on distorted images and human scores, making it dependent on the learning materials and the consistency of human opinions.

Multi-scale image quality transformer (MUSIQ) is a no-reference quality evaluation metric. The goal for this metric is to evaluate perceptual image quality without a reference image, and without changing the image resolution or aspect ratio (Ke et al., 2021). Typically, constraining factor for the performance of CNN-based models, is the fact that images are resized into a fixed shape, causing image quality degradation. Utilising a multi-scale image representation, MUSIQ is able to capture image quality at different granularities (Ke et al., 2021). The MUSIQ algorithm starts with generating a multi-scale representation from the input image that contains the original input image and its ARP resized versions. These images of different scales are then split into same size patches and further fed into the model. Due to the variety of resolution and aspect ratio within these patches, an encoding step is carried out to convert these multi-scale multi-aspect-ratio input patches into a sequence of tokens which represent the pixel, spatial and scale information (Ke et al., 2021). After the encoding step, an extra learnable CLS token is generated, and its state at the output of transformer encoder is serving as the final image representation. Finally, a fully connected layer is added on top which predicts the final image quality score (Ke et al., 2021). Due to MUSIQ only changing the input encoding, it can be used with any of the transformer variants. Using this methodology, MUSIQ can achieve a high-level end-to-end representation of any image and is able to capture distortions as well as broader image-level quality cues across different scales.

A top-down approach from semantics to distortions (TOPIQ) is an image quality assessment framework that can be utilised in both FR and NR setting (Chen et al., 2023). This approach utilises a top-down approach which involves a heuristic coarse-to-fine attention network (CFANet). The purpose for this network is to simulate the process of the human visual system by propagating semantic information in order from top to bottom. This design avoids selection from multiple features at varying scales and has proven itself effective (Chen et al., 2023). A main innovation in this process is a cross-scale attention

(CSA) mechanism. The CSA mechanism allows information propagation between levels, and it works by selecting high-level features first and they are used to guide the selection of important distortion features at lower levels. This method is inspired by transformer attention mechanism where the CSA is formulated like a query problem and is based on feature similarities, where high-level features are the queries and lower level features are formed into key-value pairs (Chen et al., 2023). The size increase of spatial feature maps makes computation expensive as the size increases quadratically from low to fine level. In the TOPIQ framework, this challenge has been addressed with a gated local pooling (GLP) block. This block reduces the size of the low-level features using gated convolution which is followed by average pooling with a pre-defined window size, which filters out redundant information and lowers the computational cost (Chen et al., 2023). In the FR setting, TOPIQ uses a distorted image and a reference image to compute feature differences, and those differences are used to locate and weight relevant regions. In the NR setting, the model can’t compute the feature differences, so it uses a learned patterns of what low-quality images typically look like. These patterns are then used in similar ways to calculate the feature differences and proceed further with the algorithm (Chen et al., 2023). This logic can be framed as the FR version measuring the difference between distorted image and reference image, and the NR version trying to figure out whether the image looks distorted or not based on learned information. Other than this difference, the FR and NR models are working with similar structure and with the same top-down logic.

Convolutional neural networks (CNN) can be used for image quality assessment, and a CNN-IQA model has been proposed by Kang *et al*. (2014). Their paper presents a model that can effectively predict image quality of distorted images with respect to human perception in a NR setting. Natural scene statistic (NSS) based approaches have shown successful results in recent years. Usually, NSS based features are extracted in transformation domain using for example, wavelet transforms or the DCT transform. These methods can be computationally expensive, and this is why in CNN-IQA, CNNs are used for learning discriminant features (Kang et al., 2014). CNN has shown effective performance, especially in object recognition tasks. CNNs can take raw images as input and include feature training within the training process, essentially learning complex feature mappings using deep structures while requiring minimal domain knowledge (Kang et al., 2014). This framework performs learning and quality prediction in local regions, which is important especially for image denoising and reconstruction problems. The framework is constructed from different stages. Firstly, a contrast normalisation is performed, after which non-overlapping patches are sampled. The CNN is then used to estimate the quality score for each patch, and an average is calculated using these patches, forming the final quality estimation for the image (Kang et al., 2014). The CNN itself is constructed from five layers, where the first layer is a convolutional layer that filters the input. The convolutional layer is followed by a pooling operation that is used to reduce each feature map to one max and one min. After that, there is two fully connected layers of 800 nodes, and last is a one-dimensional linear regression that gives the score (Kang et al., 2014). This methodology is reported to outperform methods such as CORNIA and BRISQUE, while matching results with some FR methods such as FSIM (Kang et al., 2014).

5 SAR super-resolution image quality analysis

To find the optimal image quality metric for SAR super-resolution imagery, it is first necessary to understand the applications that the imagery is used in. As SAR images can be quite complex and have many different characteristic features, there also is a variety of different things to be considered when assessing image quality. Different use-cases are prioritising different features and characteristics in many ways. The final objective for SAR super-resolution models is to make the images better for certain use. For some models the object is to reconstruct the image extremely accurately based on structure, while the goal for some other is to make the LR image more easily interpretable. Due to this, many different image quality metrics have been developed that are trying to measure image quality in their respective ways. These metrics are used to not only evaluate image quality but also used to estimate the reconstruction accuracy and performance of the used SR model through their output.

5.1 SAR Applications analysis

SAR has become an important remote sensing tool due to its ability to perform imaging independently of daylight conditions or cloud coverage. SAR is especially valuable in situations where reliable and consistent observations are required in changing environmental conditions. When compared to optical imaging methods, SAR stands out with its utilisation of microwave signals in the collection process. This allows SAR to collect scene geometry and its physical and structural properties disregarding weather conditions. Due to these capabilities, SAR is widely used in object detection, agriculture as well as in environmental monitoring. The usefulness of SAR imagery is determined by availability of data but also image quality of the produced imagery. Super-resolution is widely used image reconstruction methodology with the end-goal of achieving the optimal image quality with respect to each use-case. As there are numerous metrics, there also exists a variety of different SR models with their own areas of focus. Initial SAR image quality is typically measured using technical capabilities such as spatial resolution, radiometric consistency, noise level, sidelobe performance and phase stability (Li et al., 2022). However, to assess SR reconstructed SAR image quality, it is first necessary to determine the use-case as each application emphasises different factors in context of image quality.

An important factor to overall image quality of SAR image is the spatial resolution. Spatial resolution is important for SAR because it determines how small of a feature can be distinguished and how clearly the scene structure is reflected. Higher spatial resolution allows separation of neighbouring scatterers instead of combining them into one scatter. The separation of scatterers allows finer visibility of edges and boundaries as well as detection of fine and narrow features and patterns. The increased visibility proves itself extremely valuable especially in object detection where small details are crucial for success. Overall, low resolution reduces the amount of distinguishable information within the image, hence affecting all SAR data across different applications.

Object detection is one of the most established applications for SAR imagery. Practically object detection can be determined as identification of targets such as ships, aircrafts and other man-made structures. Object detection is challenging due to small size of some targets, and background cluttering. Ship detection is an example of object detection in maritime region. Ship detection from SAR imagery is a complex problem that is mainly revolved around separating targets from the background (Wei et al., 2020). Due to background clutters, small ships are often ignored, and dense ships are difficult to distinguish. Detail preservation is dependent on maintaining sharp target boundaries and preservation of the object structure. If these details are distorted or removed during the SR process, they can become indistinguishable or may disappear entirely into the background clutter. This is why image noise and speckle are important factors for object detection as they usually are the main factors for background complications. Strong speckle can degrade image appearance majorly and make image interpretation extremely difficult. However, speckle also contains valuable information about the scene and hence can’t be ignored completely. These factors can also cause false detections especially on small targets. The resolution determines whether a small target is constructed with enough detail that it can be separated from clutter and other image noise such as speckle (Wei et al., 2020). High resolution also allows the identification of larger objects by representing the discrete finer details that make the object recognisable in the first place. The preservation of fine details is an essential component needed for high accuracy object detection and identification. These factors are not only affecting ship detection but also can be applied to detection of aircrafts and land-based objects. Another essential contributing characteristic is the contrast between the background and the detectable target. This is practically seen as how clearly the object is standing out from the surrounding background. In ship detection, this consists of sea clutter, coastal structures and other bright scattering factors. In other words, image quality for object detection can be assessed based on the background clearness, spatial resolution and the strength of the reflection compared to the background.

SAR is also widely used in agricultural applications such as crop identification, monitoring and yield-related analysis (Liu et al., 2019). It is especially useful due to the ability of imaging regardless of weather conditions and scene illumination. This makes it suitable for repeating observations during the entirety of the growing season. SAR is majorly beneficial due to the fact that the radar signal can sense crop structure, surface roughness and dielectric properties, which means that typically agricultural classes can be identified based on the backscatter (Liu et al., 2019). For agricultural mapping, the most important characteristic is radiometric consistency. This is due to agricultural mapping often relying on backscatter comparison of images taken at different times. As agricultural analysis is based around how backscatter relates to crop properties, reliable and interpretable radar intensity information is essential (Liu et al., 2019). Also having appropriate spatial resolution is essential for monitoring purposes. The resolution of the SAR image must be fine enough to reconstruct field patterns and to highlight the field boundaries between different fields. This is another case where a super-resolution model can make a huge difference. This is especially important in areas with multiple smaller fields close by.

Sea-ice monitoring is another example of SAR imagery application. Sea ice cover is crucial part of the Arctic environment that has an important role in global climate, and has big impact on navigation as well as offshore activities in the Arctic region. (Zakhvatkina et al., 2019). Data produced with remote sensing is currently the main source of information about sea ice conditions in the Arctic. Sea-ice monitoring has similar needs for image quality than agriculture, but it has stronger emphasis on spatial structure, patterns and texture. The primary question in ice applications is whether the ice edges, leads and floe boundaries are reconstructed accurately and meaningfully without misleading details (Zakhvatkina et al., 2019). This is due to the monitoring being dependent on the ability of representing boundaries, leads and ice structures clearly for classification purposes. Classification quality is dependent on the spatial patterns and intensity captured by SAR. Separating factor between sea ice and the open water differ in the backscatter behaviour makes structure focused metrics efficient for this type of application similarly to the agricultural use-cases. Another important characteristic is texture information. Texture information carries important features such as spatial structure and surface roughness. These features are in a key position in sea ice identification and often used in combination with the pure backscatter intensity data. Spatial resolution plays also an important role in sea-ice classification as may features are small or narrow. Higher resolution enables this structure recognition as well as definition of ice boundaries and improves the visibility of local morphology. Spatial patterns are useful as they carry loads of class information as intensity levels.

5.2 SAR Metrics analysis

Simply said, image quality is dependent on the interpretability of each image. For SAR imagery the interpretation is complicated and highly dependent on the application. Despite all the complications and difficulties, the quality of an image can be evaluated based on the volume of errors and anomalies present. These types of disruptions are especially critical when considering SAR imagery which is more prone to these. Outside of these anomalies, SAR images are typically evaluated using mathematical models based on pixel differences. Most widely used metrics of this type are the MSE and the PSNR. These metrics are commonly used as the baseline for super-resolution model comparison and image quality assessment (Mohammadi et al., 2014). While providing a solid baseline, these metrics are typically thrown off by SAR characteristic speckle noise that causes two images of the same scene look significantly different on pixel level. This concentration on pixel level differences, causes metrics to be inaccurate while disregarding important structural and textural information carried by the speckle. However, the speckle noise is an inevitable part of the SAR imagery and cannot be completely removed from the images. Also, even small changes in the incidence angle of the radar can cause tremendous changes within the image itself. These factors make taking two identical SAR images almost impossible, hence having a direct impact widely on FR metrics. This is why these pixel-level metrics are typically not the most descriptive quality indicators when considering SAR imagery. For the same reason these metrics are not very practical to assess the effectiveness of SR models.

For object detection related SAR applications, FR quality assessment is often challenging due to low availability of closely matching image pairs and makes NR and RR metrics more used in many cases. This is since objects can be moving, and especially smaller targets are affected by imaging properties such as incidence angle and flight orientation of the radar itself. Producing two similar enough SAR images for FR metrics is hence difficult, and they are not so often available for object detection applications. Widely used metrics for these applications are SSIM, NIQE, BRISQUE, LPIPS and DISTS as well as learning-based NR metrics. These metrics are focused on structure, sharpness, perceptual context and image fidelity. In the training phase of SR model, these metrics can be used to evaluate model performance by comparing image quality values of the initial image against the outputted HR images values.

For agriculture related applications there is often a strong emphasis on radiometric consistency and stable spatial structure. This is due to crop monitoring techniques that are utilising imaging of the same spot over time (Liu et al., 2019). These images over time are then used for comparison of changes in backscatter or coherence across time. For this use-case, image quality should not be judged only based on the sharpness but with larger emphasis on intensity relationships that are carrying valuable crop information. For this case, metrics such as PSNR and MSE can prove to be valuable as supporting metrics for radiometric preservation while metrics like SSIM and MS-SSIM are efficient for evaluating field pattern reconstruction. Overall, structural metrics are representing image quality well due to field geometry preservation being critical. For agricultural use, pure perception-based improvements and fine spatial resolution are secondary to stable structure.

For sea-ice related applications, deep-learning and segmentation approaches to perform well as they use morphology and neighbouring structures to their advantage rather than singular pixel values (Zakhvatkina et al., 2019). With backscatter information, textural structures are important and can provide valuable insights on ice type and water region detections. The texture is especially important in situations where classes are overlapping in average backscatter but have differences in surface structure.

6 Observations

Based on the literature review, it is clear that SAR is an effective imaging technology that can provide useful resources for a variety of applications. These applications are unique by their nature, and they all have different requirements considering image quality. Spatial resolution plays a massive part for object detection purposes. With high resolution, small details are being captured efficiently, and hence object recognition becomes easier. Alongside with resolution, factors such as speckle management and target illumination are affecting image quality. High amounts of speckle and low target illumination make the object disappear into the background clutter which is typical for SAR imagery. Where object detection values high perceptibility, other different environmental applications might value stability and structural integrity. Agriculture uses SAR imagery in a completely different way than object detection-based applications. With agriculture, the end-goal is to monitor spatial patterns, which is why spatial patterns and radiometric consistency over time become more important than high resolution. This is due to crop interpretation being tied to backscatter behaviour analysis. It is also different because imagery is used over time in contrast to single image operations typically used for object detection purposes. Sea-ice monitoring follows similar principles as agriculture but with a bigger emphasis on polarisation and textural information. These factors are important for ice type and open sea detection purposes.

Based on this research, it can be observed that different existing SR image quality metrics focus on different features of SAR image quality. Methods based on pixel differences, measure numerical differences emphasising overall reconstruction accuracy but are often struggling with SAR specific features. Structural metrics compare image structures, contrast and edge preservation, which makes them sensitive to spatial consistency. Perceptual metrics assess similarity and perceived sharpness, focusing on capturing texture and structural clarity. With SAR imagery, these metrics can assess the image quality differently due to different image features influencing these metrics in different ways. Therefore, a SR model can receive different quality assessments based on the image feature that is emphasised by the metric.

7 Conclusions

Firstly, based on observations from this research, it can be concluded that no single SR image quality metric is sufficient alone for evaluating the performance of SR model or the overall quality of SAR SR image. This is caused by natural features in SAR imagery as well as image quality metrics effectively being able to evaluate only one or a few image characteristics. The characteristic features of SAR directly affect the quality of SAR SR imagery, and they cause that every image quality assessment method and metric is not effective for SAR specific SR image quality evaluation. Due to the shallow focus, SR models designed for optical imagery often have difficulties with the SAR typical image disruptions such as speckle and geometric distortions.

Secondly, observations of this research find out that requirements in context of image quality differ based on the use-case of the SAR SR imagery. Due to the different requirements, it can be concluded that the intended use-case of the SAR SR imagery must be considered when evaluating different image quality metrics. This conclusion is directly tied into the observation that different SR image quality metrics are each focused on evaluating only a specific aspect of the image.

Thirdly, observations also support the claim that instead of one single SR image quality metric, multiple different metrics should be used to evaluate the quality of SR image effectively. Multi-metric approach enables more diverse evaluation of the overall image quality and more effective assessment of the performance of a SR model regarding specific components of the reconstructed image. According to this claim, it can be suggested that these multi-metric combinations should vary depending on the SR model performance goals with respect to their application.

As the final outcome of this thesis, it can be concluded that further research is required to identify the most effective practices in terms of metric methodologies for SAR SR. In addition, future studies should explore multi-metric approaches and investigate different combinations of complementary metrics.

References

Afrasiabi, M. M., Hosseini, R., & Abbasfar, A. (2023). Combination of Single and Multi-frame Image Super-resolution: An Analytical Perspective. *JOURNAL OF LATEX CLASS FILES*, *14*(8). https://arxiv.org/abs/2303.03212v1

Ager, T., P. (2013). An Introduction to Synthetic Aperture Radar Imaging. *Oceanography*, *26*(2), 20–33. https://doi.org/10.5670/oceanog.2013.28

Ager, T., P. (2023). *The essentials of SAR: A Conceptual View of Synthetic Aperture Radar and its Remarkable Capabilities*.

Bordone Molini, A., Valsesia, D., Fracastoro, G., & Magli, E. (2020). DeepSUM: Deep Neural Network for Super-Resolution of Unregistered Multitemporal Images. *IEEE Transactions on Geoscience and Remote Sensing*, *58*(5), 3644–3656. https://doi.org/10.1109/TGRS.2019.2959248

Chen, C., Mo, J., Hou, J., Wu, H., Liao, L., Sun, W., Yan, Q., & Lin, W. (2023). TOPIQ: A Top-down Approach from Semantics to Distortions for Image Quality Assessment. *JOURNAL OF LATEX CLASS FILES*, *14*(8). https://doi.org/10.48550/arXiv.2308.03060

Dellepiane, S. G., & Angiati, E. (2014). Quality Assessment of Despeckled SAR Images. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *7*(2), 691–707. https://doi.org/10.1109/JSTARS.2013.2279501

Ding, K., Ma, K., Wang, S., & Simoncelli, E. P. (2020). Image Quality Assessment: Unifying Structure and Texture Similarity. *IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE*, *44*(5), 2567–2581. https://doi.org/10.1109/TPAMI.2020.3045810

Dong, C., Loy, C. C., He, K., & Tang, X. (2015). Image super-resolution using deep convolutional networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *38*(2), 295–307. https://doi.org/10.48550/arXiv.1501.00092

Dost, S., Saud, F., Shabbir, M., Khan, M. G., Shahid, M., & Lovstrom, B. (2022). Reduced reference image and video quality assessments: Review of methods. *EURASIP Journal on Image and Video Processing*, *2022*(1), 1. https://doi.org/10.1186/s13640-021-00578-y

Gupta, P. (2017). SpEED-QA: Spatial Efficient Entropic Differencing for Image and Video Quality. *IEEE Signal Processing Letters*, *24*(9), 1333–1337. https://doi.org/10.1109/LSP.2017.2726542

ICEYE. (2024, November 8). *Seeing the world through SAR*. https://www.iceye.com/blog/seeing-the-world-through-sar

International Telecommunication Union. (2013). *P.800.2: Mean opinion score interpretation and reporting*. https://www.itu.int/rec/T-REC-P.800.2

Irani, M., & Peleg, S. (2003). Super Resolution From Image Sequences. *ResearchGate*. https://www.researchgate.net/publication/2573175_Super_Resolution_From_Image_Sequences

Jiang, Q., Liu, Z., Gu, K., Shao, F., Zhang, X., Liu, H., & Lin, W. (2022). Single Image Super-Resolution Quality Assessment: A Real-World Dataset, Subjective Studies, and an Objective Metric. *IEEE Transactions on Image Processing*, *31*, 2279–2294. https://doi.org/10.1109/TIP.2022.3154588

Kang, L., Ye, P., Li, Y., & Doermann, D. (2014). Convolutional Neural Networks for No-Reference Image Quality Assessment. *2014 IEEE Conference on Computer Vision and Pattern Recognition*, 1733–1740. https://doi.org/10.1109/CVPR.2014.224

Karnatov, S. (2025). Analysis of PSNR, SSIM, LPIPS metrics in the context of human perception of visual similarity. *Transport Systems and Technologies*, (46). https://tst.duit.in.ua/index.php/tst/article/view/447

Ke, J., Wang, Q., Wang, Y., Milanfar, P., & Yang, F. (2021). MUSIQ: Multi-scale Image Quality Transformer. *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 5148–5157. https://doi.org/10.48550/arXiv.2108.05997

Ledig, C., Theis, L., Huszar, F., Caballero, J., Cunningham, A., Acosta, A., Aitken, A., Tejani, A., Totz, J., Wang, Z., & Shi, W. (2017). Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 4681–4690. https://doi.org/10.48550/arXiv.1609.04802

Li, S., Xu, W., & Li, Z. (2022). Review of the SBAS InSAR Time-series algorithms, applications, and challenges. *Geodesy and Geodynamics, Advances in InSAR Algorithms, Deformation Monitoring and Source Parameter Inversion*, *13*(2), 114–126. https://doi.org/10.1016/j.geog.2021.09.007

Liu, C., Chen, Z., Shao, Y., Chen, J., Hasi, T., & Pan, H. (2019). Research advances of SAR remote sensing for agriculture applications: A review. *Journal of Integrative Agriculture*, *18*(3), 506–525. https://doi.org/10.1016/S2095-3119(18)62016-7

Ma, C., Yan, B., Tan, W., & Jiang, X. (2021). Perception-Oriented Stereo Image Super-Resolution. *Proceedings of the 29th ACM International Conference on Multimedia, MM '21*, 2420–2428. https://doi.org/10.1145/3474085.3475408

Mittal, A., Moorthy, A. K., & Bovik, A. C. (2012). No-reference image quality assessment in the spatial domain. *IEEE Transactions on Image Processing: A Publication of the IEEE Signal Processing Society*, *21*(12), 4695–4708. https://doi.org/10.1109/TIP.2012.2214050

Mittal, A., Soundararajan, R., & Bovik, A. C. (2013). Making a "Completely Blind" Image Quality Analyzer. *IEEE Signal Processing Letters*, *20*(3), 209–212. https://doi.org/10.1109/LSP.2012.2227726

Mohammadi, P., Ebrahimi-Moghadam, A., & Shirani, S. (2014). Subjective and Objective Quality Assessment of Image: A Survey. *Majlesi Journal of Electrical Engineering*, *9*(1), 55–83.

Moreira, A., Prats-Iraola, P., Younis, M., Krieger, G., Hajnsek, I., & Papathanassiou, K. P. (2013). A tutorial on synthetic aperture radar. *IEEE Geoscience and Remote Sensing Magazine*, *1*(1), 6–43. https://doi.org/10.1109/MGRS.2013.2248301

Nasrollahi, K., & Moeslund, T. B. (2014). Super-resolution: A comprehensive survey. *Machine Vision and Applications*, *25*(6), 1423–1468. https://doi.org/10.1007/s00138-014-0623-4

Pickup, L. (2007). *Machine learning in multi-frame image super-resolution* [Doctoral dissertation, Oxford University, UK]. https://ora.ox.ac.uk/objects/uuid:88c6968f-1e62-4d89-bd70-604bf1f41007

Pickup, L. C., Capel, D. P., Roberts, S. J., & Zisserman, A. (2008). Bayesian Methods for Image Super-Resolution. *The Computer Journal*, *52*(1), 101–113. https://doi.org/10.1093/comjnl/bxm091

Rehman, A., & Zhou, R. (2012). Reduced-reference image quality assessment by structural similarity estimation. *IEEE Transactions on Image Processing: A Publication of the IEEE Signal Processing Society*, *21*(8), 3378–3389. https://doi.org/10.1109/TIP.2012.2197011

Ren, T., Gong, W., Bowa, V. M., Tang, H., Chen, J., & Fumeng, Z. (2021). An Improved R-Index Model for Terrain Visibility Analysis for Landslide Monitoring with InSAR. *Remote Sensing*, *13*, 1938. https://doi.org/10.3390/rs13101938

Soundararajan, R., & Bovik, A. C. (2012). RRED Indices: Reduced Reference Entropic Differencing for Image Quality Assessment. *IEEE Transactions on Image Processing*, *21*(2), 517–526. https://doi.org/10.1109/TIP.2011.2166082

Tuzova, A. A., Pavlov, V. A., Belov, A. A., & Volvenko, S. V. (2020). Comparison of Image Quality Assessment Metrics for Evaluation of Performance of Anisotropic Diffusion Filter for SAR Images. *2020 IEEE International Conference on Electrical Engineering and Photonics (EExPolytech)*, 176–179. https://doi.org/10.1109/EExPolytech50912.2020.9243957

Wang, Z., Chen, J., & Hoi, S. C. H. (2021). Deep Learning for Image Super-Resolution: A Survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *43*(10), 3365–3387. https://doi.org/10.1109/TPAMI.2020.2982166

Wei, S., Su, H., Ming, J., Wang, C., Yan, M., Kumar, D., Shi, J., & Zhang, X. (2020). Precise and Robust Ship Detection for High-Resolution SAR Imagery Based on HR-SDNet. *Remote Sensing*, *12*(1). https://doi.org/10.3390/rs12010167

Yang, J., Wright, J., Huang, T. S., & Ma, Y. (2010). Image Super-Resolution Via Sparse Representation. *IEEE Transactions on Image Processing*, *19*(11), 2861–2873. https://doi.org/10.1109/TIP.2010.2050625

Yang, W., Zhang, X., Tian, Y., Wang, W., Xue, J.-H., & Liao, Q. (2019). Deep Learning for Single Image Super-Resolution: A Brief Review. *IEEE Transactions on Multimedia*, *21*(12), 3106–3121. https://doi.org/10.1109/TMM.2019.2919431

Yue, L., Shen, H., Li, J., Yuan, Q., Zhang, H., & Zhang, L. (2016). Image super-resolution: The techniques, applications, and future. *Signal Processing*, *128*, 389–408. https://doi.org/10.1016/j.sigpro.2016.05.002

Zakhvatkina, N., Smirnov, V., & Bychkova, I. (2019). Satellite SAR Data-based Sea Ice Classification: An Overview. *Geosciences*, *9*(4). https://doi.org/10.3390/geosciences9040152

Zhang, R., Isola, P., Efros, A. A., Shechtman, E., & Wang, O. (2018). The Unreasonable Effectiveness of Deep Features as a Perceptual Metric. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 586–595. https://doi.org/10.48550/arXiv.1801.03924