Paradigma, Vol. 27, No. 1, March 2025, P-ISSN 1410-5063, E-ISSN: 2579-3500, Page. 21-29
Published: LPPM Universitas Bina Sarana Informatika

PARADIGMA

website: http://jurnal.bsi.ac.id/index.php/paradigma/

Efficient Image Transmission for Autonomous Systems Using Residual Dense Feature Networks Over LoRa Networks

Muhamad Fadly Rizqy Praptawilaga[1], Galura Muhammad Suranegara[2], Arief Suryadi Satyawan[3]

[1,2] Department of Telecommunication System, Universitas Pendidikan Indonesia, Purwakarta, Indonesia
[3] National Research and Innovation Agency, Bandung, Indonesia

ARTICLE INFORMATION

Artikel History:
Received: December 4, 2024
Revised: January 10, 2025
Accepted: February 25, 2025
Available Online: March 6, 2025

Keyword:

Autonomous Systems
LoRa communication
Image processing
Image transmission
Super resolution

ABSTRACT

Autonomous systems face challenges in transmitting high-quality images over bandwidth-constrained networks like LoRa, which operates at data rates of 0.3–50 kbps. This study proposes the Residual Dense Feature Network (RDF Net), a super-resolution model designed to optimize image transmission within the constraints of LoRa networks. By leveraging Contrast-Aware Channel Attention (CCA), Enhanced Spatial Attention (ESA), Blueprint Separable Convolution (BSConv), and a progressive approach, RDF Net achieves 20x upscaling, enabling low-resolution images (40x40 pixels) to be reconstructed into high-resolution outputs (800x800 pixels) on a central server. Experimental evaluations demonstrate that Model-4, combining CCA and ESA, delivers state-of-the-art perceptual quality and structural fidelity, while Model-3, using ESA, offers a computationally efficient alternative for resource-constrained scenarios. Simulations of LoRa's bandwidth limitations reveal that transmitting a single 40x40 image requires approximately 0.208–0.56 seconds at a data rate of 50 kbps. While this demonstrates the feasibility of near real-time communication, the trade-off between latency and visual fidelity remains a critical consideration, particularly for latency-sensitive applications. These findings underscore RDF Net's potential to address the challenges of high-quality visual communication in bandwidth-constrained environments, paving the way for enhanced autonomous system applications. Further optimization, including adaptive compression strategies, and testing on actual LoRa hardware are recommended to validate its performance in real-world scenarios and explore its applicability to diverse autonomous systems.

Corresponding Author:

Muhamad Fadly Rizqy Praptawilaga,
Department of Telecommunication System,
Universitas Pendidikan Indonesia,
Purwakarta, West Java, Indonesia, 41115,
Email: fadlyrizqy@upi.edu

INTRODUCTION

Autonomous systems in industrial, campus, and logistics environments often operate under constrained energy conditions, requiring communication systems that are simple and energy efficient. LoRa (Long Range) is a widely used communication technology in such scenarios due to its low power consumption and long-range capabilities. However, LoRa is limited by a narrow bandwidth range of 0.3–50 kbps data rates, making it challenging to transmit large datasets, such as images (Dede et al., 2024; Murray et al., 2021; Pham et al., 2023).

This limitation leads to longer transmission times or reduced image quality, which is problematic for autonomous systems that rely on high-quality visual data. To address this, efficient image transmission methods are essential. This study proposes capturing low-resolution images (40x40

DOI: https://doi.org/10.31294/paradigma.v27i1.7584

pixels) that fit within LoRa's bandwidth constraints and using advanced super-resolution techniques to upscale them to 800x800 pixels on a central server, ensuring both transmission efficiency and image quality.

Previous SR models, such as Generative Adversarial Networks (GANs) (Park & Lee, 2021; Shang et al., 2020), and Attention based Back-Projection Networks (Qin et al., 2024, p. 3) have primarily focused on upscaling by factors of 16x. However, these methods are not optimized for the specific challenges of autonomous system imagery and bandwidth limitations. Non-GAN approaches, such as those leveraging residual architectures, offer a more computationally efficient alternative, but they have not been widely explored for extreme upscaling tasks, particularly in bandwidth-constrained autonomous systems.

This study introduces the Residual Dense Feature Network (RDF Net), an innovative architecture designed for 20x upscaling that incorporates advanced attention mechanisms, such as Contrast-Aware Channel Attention (CCA) and Enhanced Spatial Attention (ESA). RDF Net addresses the unique challenges of autonomous systems by improving feature extraction and image reconstruction while maintaining computational efficiency. Unlike prior approaches, RDF Net's architecture is specifically optimized for low-resolution image inputs under stringent bandwidth constraints.

The core contribution of this research lies in its integration of both CCA and ESA mechanisms, which enhance image quality by refining critical features at both the channel and spatial levels. RDF Net builds upon the Residual Local Feature Network (F. Kong et al., 2022) by adopting dense connections that enable the network to capture a broader range of features essential for high-quality image reconstruction. Additionally, RDF Net employs a progressive upscaling approach inspired by (Chudasama et al., 2022), allowing the model to reconstruct images step-by-step for greater precision and efficiency. Optimized convolutional layers, specifically Blueprint Separable Convolutions (BSConv), further enhance computational efficiency without sacrificing performance (Z. Li et al., 2022).

By addressing the challenges of transmitting high-quality visual data over bandwidth-constrained networks, this study provides a significant step forward in enabling reliable and efficient communication for autonomous systems. It introduces RDF Net, an SR architecture for extreme image super-resolution tailored to the bandwidth limitations of LoRa-based autonomous systems. By integrating advanced attention mechanisms (CCA and ESA) and leveraging state-of-the-art design choices such as progressive upscaling and BSConv layers, RDF Net achieves a balance between computational efficiency and visual fidelity. This work demonstrates state-of-the-art performance across multiple datasets, underscoring its potential for practical applications in low-latency and bandwidth-limited scenarios. The findings also highlight RDF Net's novelty compared to existing SR models, offering efficient solution for extreme upscaling in bandwidth-limited environments. Future research will focus on further optimizing RDF Net and exploring alternative compression methods to enhance real-time performance in broader autonomous system environments.

RESEARCH METHOD

1. Proposed Model

This study introduces the Residual Dense Feature Network (RDF Net), an advanced architecture designed to perform progressive super-resolution (SR), aimed at transforming low-resolution (LR) inputs into high-resolution (HR) outputs with an impressive 20× upscaling factor. As depicted in Figure 1, the RDF Net operates in a structured two-branch configuration, ensuring high-quality restoration and precise upscaling of image details at multiple scales.

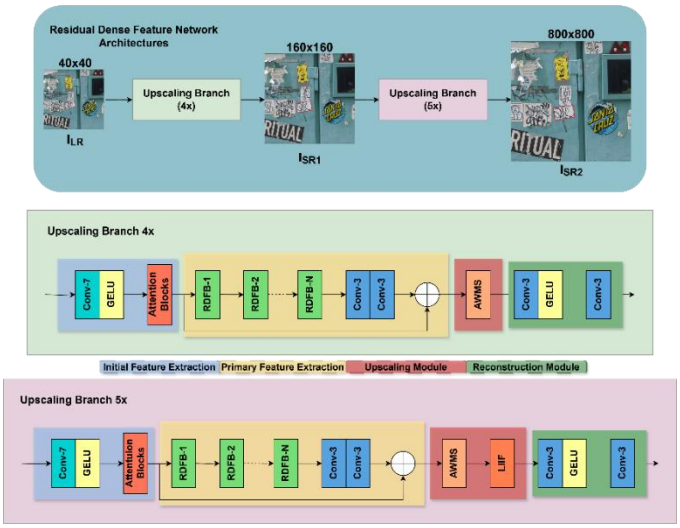

Figure 1. Diagram of Proposed RDF Network Model

The first branch of RDF Net begins by applying a 4× upscaling to the low-resolution input

image I_{LR}, generating an intermediate super-resolved image I_{SR1}. This initial result is then passed to the second branch, where it undergoes an additional 5× upscaling, resulting in the final high-resolution output I_{SR2}. This progressive upscaling strategy leverages the combined strength of several key modules, ensuring both computational efficiency and high-quality restoration across diverse scales (Chudasama & Upla, 2020; A. Liu et al., 2022; Y. Wang et al., 2023).

Each branch is comprised of four essential modules: Initial Feature Extraction (IFE), Primary Feature Extraction (PFE), Upscaling, and Reconstruction. These modules work sequentially, processing the image step-by-step to maintain a cohesive flow and precise enhancement of the features.

In the Initial Feature Extraction (IFE) module, essential foundational features from the LR image are captured using a 7×7 convolutional kernel. This large kernel allows the model to gather broad contextual information, which is essential for high-resolution reconstruction. Additionally, the Gaussian Error Linear Unit (GELU) activation function is used to facilitate nuanced learning (Liang et al., 2021; Z. Liu et al., 2021), while Attention Blocks are incorporated to refine and emphasize the most important details within the image. These components work together to establish a strong foundation, producing feature maps that retain crucial structural elements of the original low-resolution image.

Building on the IFE, the Primary Feature Extraction (PFE) module enhances these foundational features by utilizing dense connectivity patterns based (Ju et al., 2023; G. Li & Zhu, 2022), and Residual Local Feature Blocks (RLFBs) structures, to create Residual Dense Feature Block (RDFB). Each RDFB within PFE employs Blueprint-separable Convolution (BSConv) layers, as illustrated in Figure 2, which shows the Diagram of RDFB and BSConv.

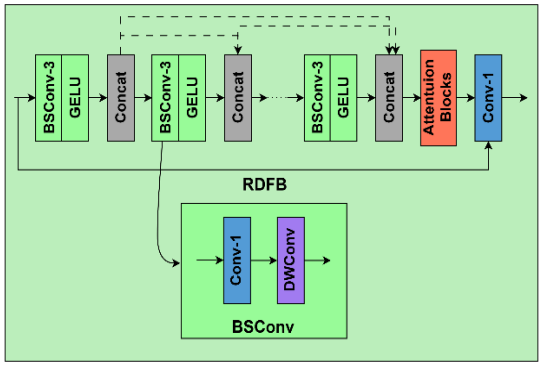

Figure 2. Diagram of RDFB and BSConv

BSConv consist of a combination of 1×1 pointwise and 3×3 depth-wise convolutions, which optimize computational efficiency while maintaining high feature extraction accuracy (Bai & Liang, 2024; Haase & Amthor, 2020; Z. Li et al., 2022; Zhang et al., 2022). Attention mechanisms further refine local features, and long skip connections address the vanishing gradient problem, supporting detailed feature propagation. This module ensures that the feature map is enriched and ready for the final stages of upscaling and reconstruction.

The next phase involves the Upscaling Module, which increases the resolution of the image in two distinct phases. In the first phase, the image is upscaled by 4× using the Adaptive Weight Multi-Scale Reconstruction (AWMS) module. This module consists of three parallel convolutional branches with kernel sizes of 3×3, 5×5, and 7×7, followed by a Pixel Shuffle operation to refine the upscaled image (Z. Li et al., 2021). The second phase utilizes the Local Implicit Image Function (LIIF), which employs grid sampling and feature interpolation to further upscale the image by an additional 1.25×, achieving the desired 5× upscaling. The combined use of AWMS and LIIF ensures that the image quality is preserved while achieving precise upscaling, even for non-integer scaling factors (Chen et al., 2020). The architecture of these modules is clearly depicted in Figure 3, showcasing the complementary roles of AWMS and LIIF in enhancing the image resolution.

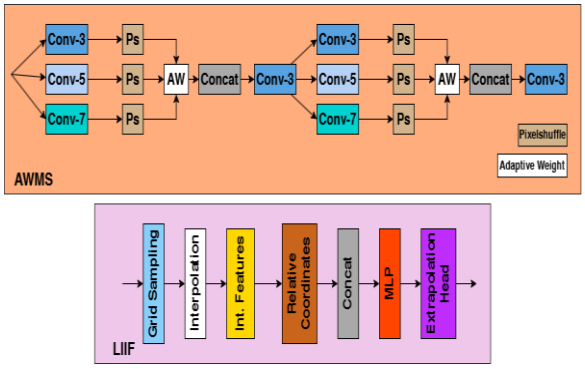

Figure 3. Diagram of AWMS and LIIF Module

Finally, the Reconstruction Module processes the enhanced feature maps from the Upscaling Module to produce the SR output. This module employs two 3×3 convolutional layers to consolidate and refine the final image. This final convolution transforms the upscaled features into a clear, high-resolution output, completing the SR process and ensuring that the output image retains sharpness and structural fidelity.

2. Model Configurations

To evaluate the effectiveness of attention mechanisms in the RDF Net architecture, we experiment with various configurations within the Initial Feature Extraction (IFE) module and Residual Dense Feature Blocks. Specifically, we integrate Contrast-Aware Channel Attention (CCA) and Enhanced Spatial Attention (ESA) modules, each tailored to enhance different aspects of feature extraction.

The CCA module amplifies critical features across channels to increase contrast within feature maps using sigmoid activation, while the ESA module focuses on spatial attention by processing spatial information through convolution, max pooling, and upsampling layers. Figure 4 illustrates the structure and operational flow of these attention blocks, detailing the architecture of each module.

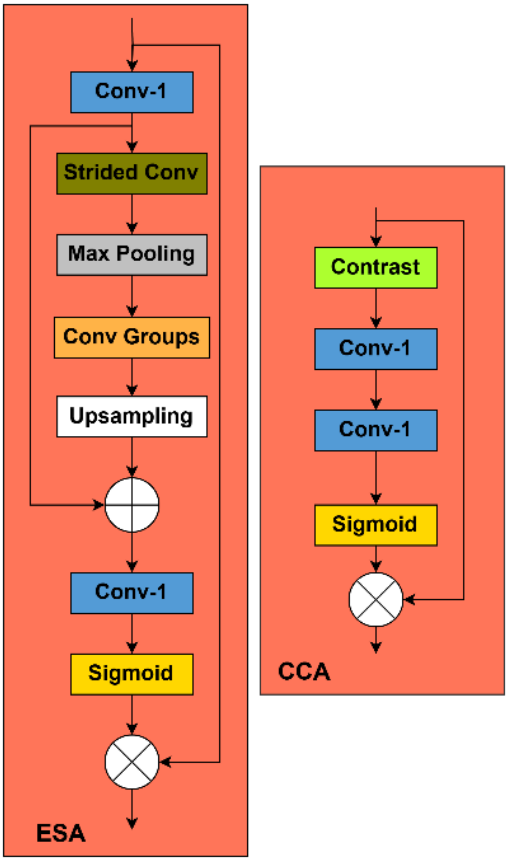

Figure 4. Diagram of CCA and ESA Block

In our experiments, we assess four model variants to determine the impact of each attention configuration, as outlined in Table 1.

Table 1. Configuration of each model

No. Model	Attention type	Description
1	None	Baseline Model
2	CCA	Add CCA
3	ESA	Add ESA
4	CCA+ESA	Combine CCA+ESA

Model 1 serves as the baseline without any attention mechanisms in IFE and RDFB, providing a control for comparison. Model 2 introduces the CCA module, designed to adjust contrast and selectively enhance critical channels for improved focus on essential information. Model 3 incorporates the ESA module, which captures spatial context and fine-grained spatial details across the image. Finally, Model 4 combines both CCA and ESA modules, integrating spatial and channel attention to enrich feature representation comprehensively. This systematic evaluation enables us to identify the optimal attention configuration for achieving high-resolution detail restoration, offering valuable insights for further refinement of super-resolution models in bandwidth-constrained environments.

3. Experiment

The proposed model was implemented using PyTorch and tested on a curated dataset, OurDatasets, consisting of 593 images. High-resolution (HR) patches of size 800×800 pixels were generated via random cropping, resulting in three distinct crops per image. The dataset was refined using the Laplacian activation variance method, filtering out patches with variance below 100, leaving 1,500 high-quality patches. These were divided into 1,180 images for training, 295 for validation, and 25 for testing.

For performance evaluation, RDF Net was tested on the OurDatasets test set (25 images) and external datasets, including DIV8K and DIV2K, each containing 100 high-resolution images. This provided a comprehensive evaluation of the model's generalization and reconstruction capabilities (Baghel et al., 2023; X. Kong et al., 2021; Zhong Xueliangand Luo, 2023). Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index Measure (SSIM) were used as performance metrics to assess pixel-level accuracy and perceptual quality, respectively.

To maintain consistency across model configurations, each model was trained with three residual layers and eight residual blocks. The model was trained using Adam optimizer with a learning rate of 10^-3, which was halved every 10 epochs. The Adam parameters were set to $\beta1 = 0.9$, $\beta2 = 0.999$, and $\varepsilon = 10^{-8}$ to ensure stable updates. The training employed a feature-based loss function, inspired by recent work (Andonian et al., 2021; F. Kong et al., 2022), utilizing a lightweight "Convk3s1-Tanh-Convk3s1" architecture for feature learning, reducing computational load compared to pre-trained models like VGG-19. The model was trained for 100 epochs with a batch size of 1, given memory constraints. Training was carried out on an NVIDIA RTX 3080 Ti GPU with 12GB VRAM, ensuring efficient training and evaluation.

The evaluation process incorporated statistical methods, particularly comparing the PSNR and SSIM values across datasets. The use of these metrics ensured an objective measure of the model's performance across various high-resolution content types, validating both pixel accuracy and visual quality.

RESULTS AND DISCUSSION

1. Model Training Performance and Computational Efficiency

The training loss trajectories of the four model configurations over 100 epochs provide valuable insights into their performance, as illustrated in Figure 5. Model-1 demonstrates the fastest convergence, characterized by a sharp drop in training loss during the first 20 epochs and stabilizing at the lowest final training loss of 0.01803 by Epoch 100. In contrast, Model-4 begins with the highest initial loss but converges gradually, stabilizing at 0.01818. This gradual convergence suggests better generalization capabilities compared to the efficiency of Model-1. Models 2 and 3 exhibit intermediate trajectories, with Model-3 requiring slightly longer to converge. This delay is attributed to its use of Enhanced Spatial Attention (ESA), reflected in the less steep slope of its

curve during the initial 50 epochs.

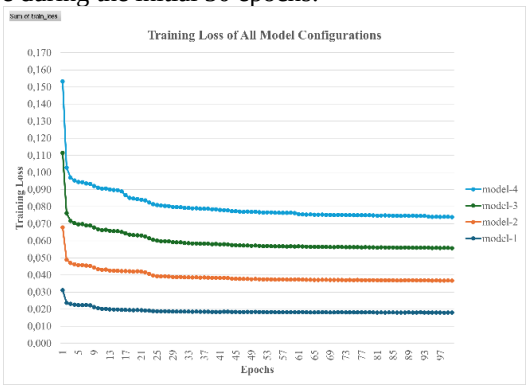

Figure 5. Training loss across 100 epochs for all model configurations

Overall, the trends in training loss graph reflect the varying convergence behaviors of each model. Model-1 emerges as the most efficient in training loss reduction, while Model-4 balances slower convergence with enhanced generalization performance.

For Validation loss, Figure 6 illustrates the validation loss trajectories of the four model configurations across 100 epochs. All models exhibit a steady reduction in validation loss, with a sharp decrease during the initial 20 epochs, followed by a more gradual stabilization as training progresses. Model-1 achieves the lowest final validation loss of 0.01796 by Epoch 100, highlighting its superior generalization capabilities.

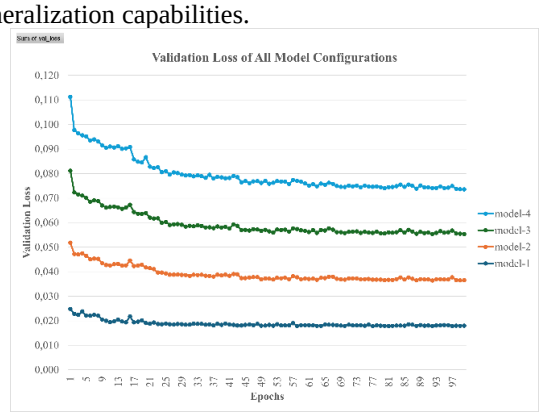

Figure 6. Validation loss across 100 epochs for all model configurations

Models 2 and 3 display slightly slower convergence, with final validation losses of 0.01861 and 0.01878, respectively, reflecting the impact of incorporating attention mechanisms. Model-2, which includes Contrast-Aware Channel Attention (CCA), shows consistent performance across epochs, as indicated by its relatively smooth trajectory. In comparison, Model-3, which utilizes Enhanced Spatial Attention (ESA), demonstrates a slightly more gradual decline in loss during the first 50 epochs, suggesting the added complexity of spatial feature extraction.

Model-4 begins with the highest initial validation loss of 0.04356, consistent with its training loss trend, but it gradually converges to a final loss of 0.01821 by Epoch 100. This trajectory highlights its robust generalization capabilities despite slower initial learning dynamics. The differences in loss trajectories across the models underscore the trade-off between convergence speed and feature extraction complexity.

Regarding computational efficiency, as illustrated in Figures 7, presents the total training duration for each model configuration, highlighting the trade-off between computational efficiency and model complexity. Model-1, which does not incorporate attention mechanisms, completes training in the shortest time of 7.44 hours, reflecting its simplicity and lower computational demand.

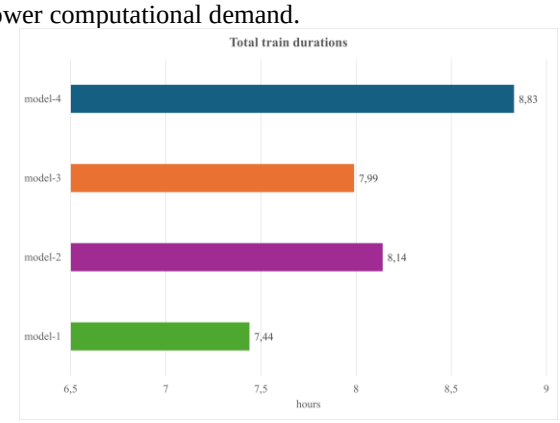

Figure 7. Total Training Duration

Model-2 and Model-3, which include CCA and ESA, respectively, require slightly longer durations of 8.14 hours and 7.99 hours. The additional time for Model-2 suggests that channel-based attention mechanisms increase computational load, whereas Model-3's use of spatial attention results in only a marginal increase in training time.

Model-4, which combines both CCA and ESA, requires the longest training duration of 8.83 hours. This reflects the added complexity of integrating dual attention mechanisms, which enhance feature extraction and image restoration performance but at the cost of increased computational demands. The results indicate that more complex models like Model-4 prioritize feature richness and generalization over computational efficiency, making them suitable for high-performance applications where training time is less critical.

2. Analysis of Model Configuration

The RDF Net model was tested using simulated datasets (Ourdatasets) collected from environments where autonomous systems typically operate, as well as the widely used DIV2K and DIV8K datasets, to evaluate the model's generalizability. While these datasets provide a robust basis for benchmarking, it is important to note that results may vary when tested on real-world data or in real-time applications due to factors such as noise, environmental variability, and hardware limitations.

The performance of attention mechanisms in enhancing image restoration is demonstrated in Table 2, which compares results across three datasets: DIV2K Val, OurDatasets, and DIV8K Test. Model-1, the baseline without attention mechanisms, serves as

the reference, showing solid results with a PSNR of 20.4977 and SSIM of 0.5831 on DIV2K Val, 19.7949 PSNR and 0.5929 SSIM on OurDatasets, and 23.3077 PSNR with 0.6474 SSIM on DIV8K Test.

Integrating Contrast-Aware Channel Attention (CCA) in Model-2 results in modest improvements over the baseline. Although the increases in PSNR and SSIM are slight, the most noticeable change is in the SSIM for the DIV8K Test dataset, where the value reaches 0.6478. This demonstrates that CCA has a moderate yet consistent positive impact on image restoration quality.

Model-3, which utilizes Enhanced Spatial Attention (ESA), shows more substantial performance improvements. This model achieves the highest PSNR among single-attention models on the DIV2K Val dataset (20.6050 PSNR, 0.5848 SSIM), with notable improvements in PSNR on OurDatasets (19.8992 PSNR) and a slight increase in SSIM (0.5948). ESA excels in capturing spatial features, improving the perceptual accuracy of fine details in the restored images.

The most significant improvements are observed in Model-4, which combines both CCA and ESA attention mechanisms. This model delivers the highest performance across all datasets, with the best PSNR (20.6250) and SSIM (0.5863) on DIV2K Val and the highest SSIM (0.6489) on DIV8K Test. The combination of these attention mechanisms enhances both structural similarity and perceptual quality, making Model-4 the most effective for image restoration across multiple high-resolution datasets. The results suggest that combining attention mechanisms improves both fine detail restoration and overall image quality.

Table 2. Average PSNR and SSIM results of each model configurations

No. Model	Configurations	DIV2K Val		OurDatasets		DIV8K Test	
	Attention Type	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
1	None	20,4977	0,5831	19,7949	0,5929	23,3077	0,6474
2	CCA	20,5702	0,5841	19,8260	0,5920	23,4430	0,6478
3	ESA	20,6050	0,5848	19,8992	0,5948	23,4484	0,6479
4	ESA+CCA	20,6250	0,5863	19,8775	0,5945	23,5162	0,6489

3. Comparison Results of Model Predictions

At the time of this study, there were no existing models capable of performing 20x super-resolution on-the-go under bandwidth-constrained environments. Therefore, the comparison was conducted using a controlled scenario prepared in this research, where multiple model configurations (Model 1 to Model 4) from the proposed RDF Net were evaluated to benchmark performance. These configurations were tested across multiple datasets, focusing on metrics such as Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index Measure (SSIM), to assess distortion reduction and structural fidelity.

The performance of the super-resolution models is evaluated across multiple datasets, focusing on distortion reduction and structural fidelity. Table 3 provides a quantitative comparison of the models' performance in terms of PSNR and SSIM for various frames. On the DIV2K Val dataset, Model 1 (the baseline) achieves a PSNR of 20.57 and an SSIM of 0.5815 for the 0834.png frame. Models 2 and 3 show slight improvements, with Model 3 reaching a PSNR of 20.59 and an SSIM of 0.5829. Model 4, which incorporates both Contrast-Aware Channel Attention (CCA) and Enhanced Spatial Attention (ESA), outperforms all others, achieving a PSNR of 20.63 and an SSIM of 0.5830. For the 0900.png frame, Model 4 again leads with a PSNR of 17.55 and SSIM of 0.5042, showing its effectiveness across various frame types.

Table 3. Quantitative Result PSNR and SSIM results of Single Frame

Frame	Metric	Model 1	Model 2	Model 3	Model 4
0834.png	PSNR	20.57	20.53	20.59	20.63
	SSIM	0.5815	0.5813	0.5829	0.5830
0900.png	PSNR	17.45	17.43	17.51	17.55
	SSIM	0.4962	0.4952	0.5014	0.5042
Frame_01.png	PSNR	17.49	17.48	17.49	17.47
	SSIM	0.4285	0.4267	0.4289	0.4295
Frame_60.png	PSNR	19.25	19.26	19.28	19.27
	SSIM	0.5842	0.5823	0.5839	0.5851
1495.png	PSNR	30.40	30.30	30.40	30.43
	SSIM	0.8495	0.8494	0.8501	0.8499
1455.png	PSNR	23.55	23.68	23.46	23.62
	SSIM	0.7034	0.7055	0.7033	0.7063

In OurDatasets, Model 4 continues to show the best performance. For Frame_01.png, it reaches

the highest SSIM of 0.4295, while on Frame_60.png, it strikes a balance with a PSNR of 19.27 and an SSIM of 0.5851. These results emphasize Model 4's ability to preserve structural integrity even in challenging frames with complex textures and fine details. On the DIV8K Test dataset, Model 4 leads with the best PSNR and SSIM for 1495.png, achieving a PSNR of 30.43 and SSIM of 0.8499. It also delivers strong performance on the 1455.png frame with a PSNR of 23.62 and an SSIM of 0.7063, showcasing its ability to reduce distortion and preserve fine details.

Visual evaluations, as shown in Figure 8, align with these quantitative results. On the DIV2K Val dataset, Model 4 produces outputs closest to the ground truth, effectively preserving texture details and improving edge clarity. This is especially evident in frames like 0834.png and 0900.png, where Model 4 manages to reduce noise and maintain fine details, outshining the other models. In OurDatasets, Model 4's output for Frame_01.png and Frame_60.png preserves structural details, such as building edges and shadows, more effectively than the other models. The DIV8K Test dataset further confirms Model 4's superiority in capturing intricate textures and edges, as seen in frames like 1495.png and 1455.png, where the model minimizes artifacts and aligns closely with the ground truth.

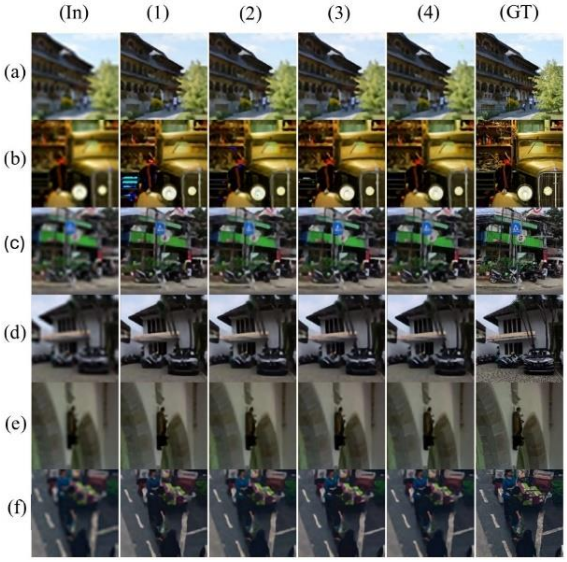

Figure 8. Visual comparison of super-resolution results across different model configurations on selected frames from each dataset, displayed at 10% scale. Images (a) 0834.png and (b) 0900.png are from the DIV2K Val dataset, (c) Frame_01.png and (d) Frame_60.png are from OurDatasets, and (e) 1495.png and (f) 1455.png are from the DIV8K Test dataset. Each model configuration labeled (1)-(4), input (In) and the ground truth (GT).

The overall performance analysis across both quantitative metrics and visual outputs consistently demonstrates that Model 4 is the top performer. Its combination of CCA and ESA attention mechanisms gives it a significant advantage in reducing distortion while maintaining structural fidelity and clarity. Although Model 3 offers a strong alternative for scenarios where PSNR is prioritized, Model 4 remains the most balanced and adaptable choice for high-quality super-resolution applications. Models 1 and 2, while effective as baselines, do not match the superior performance of Model 4, especially in terms of structural preservation and fine detail recovery.

The findings of this study have significant practical implications for real-world autonomous systems, such as remote sensing operating in bandwidth-constrained environments. RDF Net's ability to balance computational efficiency and visual fidelity makes it a promising candidate for real-time image communication in such applications.

Despite the absence of prior studies addressing 20x upscaling for bandwidth-limited environments, RDF Net builds upon existing approaches like GANs and residual networks by introducing novel attention mechanisms and progressive upscaling techniques. These innovations position RDF Net as a critical advancement in extreme super-resolution, filling a significant gap in the current literature while offering new opportunities for future exploration.

CONCLUSION

This study has successfully demonstrated the capabilities of the Residual Dense Feature Network (RDF Net) for addressing the critical challenge of image super-resolution in bandwidth-constrained environments, such as LoRa networks. The proposed RDF Net achieves an impressive 20x upscaling, transforming low-resolution 40x40 pixel images into high-quality 800x800 pixel outputs. Through the integration of advanced attention mechanisms, namely Contrast-Aware Channel Attention (CCA) and Enhanced Spatial Attention (ESA), the model achieves state-of-the-art results in both perceptual quality and structural fidelity across diverse datasets.

The research highlights two key contributions. Model-4, which integrates Contrast-Aware Channel Attention (CCA) and Enhanced Spatial Attention (ESA), achieves superior image restoration with the highest PSNR of 20.6250 and SSIM of 0.5863 on the DIV2K validation dataset and the best SSIM of 0.6489 on the DIV8K test dataset, making it optimal for scenarios demanding high precision. Meanwhile, Model-3, utilizing only ESA, offers a resource-efficient alternative with competitive performance, achieving a PSNR of 20.6050 and SSIM of 0.5848 on DIV2K and a PSNR of 19.8992 and SSIM of 0.5948 on OurDatasets, demonstrating its suitability for constrained resource environments while maintaining robust image quality.

These findings directly address the problem of balancing image quality and transmission efficiency over LoRa networks, demonstrating RDF Net's potential in enhancing autonomous system applications. Despite these advancements, challenges

persist in real-time image transmission over LoRa, where broadcasting a single 40x40 image still requires approximately 0.208–0.56 seconds at a data rate of 50 kbps. This delay underscores the need for further optimization of RDF Net, exploration of alternative compression methods, and testing on real-world LoRa environments.

ACKNOWLEDGMENT

We extend our special acknowledgment to the Indonesia Endowment Fund for Education (Lembaga Pengelola Dana Pendidikan) for supporting this study through the National Research and Innovation Agency's Riset dan Inovasi untuk Indonesia Maju (RIIM) program (BRIN).

REFERENCES

Andonian, A., Park, T., Russell, B. C., Isola, P., Zhu, J.-Y., & Zhang, R. (2021). Contrastive Feature Loss for Image Prediction. *2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW)*, 1934–1943. https://api.semanticscholar.org/CorpusID:244117332

Baghel, N., Dubey, S. R., & Singh, S. K. (2023). PTSR: Patch Translator for Image Super-Resolution. *ArXiv*, *abs/2310.13216*. https://api.semanticscholar.org/CorpusID:264405985

Bai, H., & Liang, X. (2024). A very lightweight image super-resolution network. *Scientific Reports*, *14*(1), 13850. https://doi.org/10.1038/s41598-024-64724-y

Chen, Y., Liu, S., & Wang, X. (2020). Learning Continuous Image Representation with Local Implicit Image Function. *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 8624–8634. https://api.semanticscholar.org/CorpusID:229221619

Chudasama, V., & Upla, K. (2020). E-ProSRNet: An enhanced progressive single image super-resolution approach. *Computer Vision and Image Understanding*, *200*, 103038. https://doi.org/10.1016/j.cviu.2020.103038

Chudasama, V., Upla, K., Raja, K., Ramachandra, R., & Busch, C. (2022). Compact and progressive network for enhanced single image super-resolution—ComPrESRNet. *The Visual Computer*, *38*(11), 3643–3665. https://doi.org/10.1007/s00371-021-02193-4

Dede, O. L., Jalajamony, H. M., & Fernandez, R. E. (2024). Image Transmission over LoRa-Based Networks: A Performance Study Using Image Compression and Reconstruction Methods. *2024 IEEE 21st Consumer Communications & Networking Conference (CCNC)*, 642–643. https://doi.org/10.1109/CCNC51664.2024.10454687

Haase, D., & Amthor, M. (2020). Rethinking Depthwise Separable Convolutions: How Intra-Kernel Correlations Lead to Improved MobileNets. *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 14588–14597. https://doi.org/10.1109/CVPR42600.2020.01461

Ju, R.-Y., Chen, C.-C., Chiang, J.-S., Lin, Y.-S., Chen, W.-H., & Chien, C.-T. (2023). *Resolution Enhancement Processing on Low Quality Images Using Swin Transformer Based on Interval Dense Connection Strategy*. http://arxiv.org/abs/2303.09190

Kong, F., Li, M., Liu, S., Liu, D., He, J., Bai, Y., Chen, F., & Fu, L. (2022). Residual Local Feature Network for Efficient Super-Resolution. *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 765–775. https://doi.org/10.1109/CVPRW56347.2022.00092

Kong, X., Zhao, H., Qiao, Y., & Dong, C. (2021). ClassSR: A General Framework to Accelerate Super-Resolution Networks by Data Characteristic. *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 12011–12020. https://doi.org/10.1109/CVPR46437.2021.01184

Li, G., & Zhu, Y. (2022). A Novel Dual Dense Connection Network for Video Super-resolution. *CoRR*, *abs/2203.02723*. https://doi.org/10.48550/ARXIV.2203.02723

Li, Z., Liu, Y., Chen, X., Cai, H., Gu, J., Qiao, Y., & Dong, C. (2022). Blueprint Separable Residual Network for Efficient Image Super-Resolution. *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 832–842. https://doi.org/10.1109/CVPRW56347.2022.00099

Li, Z., Wang, C., Wang, J., Ying, S., & Shi, J. (2021). Lightweight adaptive weighted network for single image super-resolution. *Computer Vision and Image Understanding*, *211*, 103254. https://doi.org/https://doi.org/10.1016/j.cviu.2021.103254

Liang, J., Cao, J., Sun, G., Zhang, K., Van Gool, L., & Timofte, R. (2021). SwinIR: Image Restoration Using Swin Transformer. *2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW)*, 1833–1844. https://doi.org/10.1109/ICCVW54120.2021.00210

Liu, A., Li, S., & Chang, Y. (2022). Image super-resolution using progressive residual multi-dilated aggregation network. *Signal, Image and Video Processing*, *16*(5), 1271–1279. https://doi.org/10.1007/s11760-021-02078-y

Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Lin, S., & Guo, B. (2021). Swin Transformer: Hierarchical Vision Transformer using Shifted Windows. *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, 9992–10002. https://doi.org/10.1109/ICCV48922.2021.00986

Murray, D., Koziniec, T., & deSouza, A. (2021). Exploring the Characteristics and Limits of LoRaWAN. *2021 IEEE International Conference on Internet of Things and Intelligence Systems (IoTaIS)*, 110–116. https://doi.org/10.1109/IoTaIS53735.2021.9628831

Park, S., & Lee, E. (2021). One-to-many Approach for Improving Super-Resolution. *ArXiv*, *abs/2106.10437*. https://api.semanticscholar.org/CorpusID:235490650

Pham, V. D., Vishnevsky, V., Nguyen, D. C., & Kirichek, R. (2023). *LoRa Mesh Network for Image Transmission: An Experimental Study* (Y. Koucheryavy & A. Aziz, Eds.; pp. 606–617). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-30258-9_54

Qin, Y., Wang, J., Cao, S., Zhu, M., Sun, J., Hao, Z., & Jiang, X. (2024). SRBPSwin: Single-Image Super-Resolution for Remote Sensing Images Using a Global Residual Multi-Attention Hybrid Back-Projection Network Based on the Swin Transformer. *Remote Sensing*, *16*(12). https://doi.org/10.3390/rs16122252

Shang, T., Dai, Q., Zhu, S., Yang, T., & Guo, Y. (2020). Perceptual Extreme Super Resolution Network with Receptive Field Block. *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 1778–1787. https://doi.org/10.1109/CVPRW50498.2020.00228

Wang, Y., Liu, W., Sun, W., Meng, X., Yang, G., & Ren, K. (2023). A Progressive Feature Enhancement Deep Network for Large-Scale Remote Sensing Image Superresolution. *IEEE Transactions on Geoscience and Remote Sensing*, *61*, 1–13. https://doi.org/10.1109/TGRS.2023.3310518

Zhang, G., Weng, H., Liu, R., Zhang, M., & Zhang, Z. (2022). Point Clouds Classification of Large Scenes based on Blueprint Separation Convolutional Neural Network. *2022 IEEE 25th International Conference on Computer Supported Cooperative Work in Design, CSCWD 2022*, 525–530. https://doi.org/10.1109/CSCWD54268.2022.9776241

Zhong Xueliang and Luo, J. (2023). Classification-Based and Lightweight Networks for Fast Image Super Resolution. In A. and A. P. and J. C. Iliadis Lazaros and Papaleonidas (Ed.), *Artificial Neural Networks and Machine Learning – ICANN 2023* (pp. 140–152). Springer Nature Switzerland.