Vol. 8, No. 2, April 2026

Journal of Information Systems and Informatics

Published By Asosiasi Doktor Sistem Informasi Indonesia

Reinforcement Learning–Guided Hyperparameter Tuning for U-Net-Based Super-Resolution of Brain MRI Under Synthetic Degradation

Suci Ramadini[1], Julian Supardi[2]

[12]Informatics Department, Postgraduate Program, Sriwijaya University, Palembang, Indonesia

Received:
December 8, 2025
Revised:
March 11, 2026
Accepted:
March 27 2026
Published:
April 12, 2026

Corresponding Author:
Author Name*:
Suci Ramadini
Email*:
09012682327012@student.unsri.ac.id

DOI:
10.63158/journalisi.v8i2.1565

Abstract. Low-resolution magnetic resonance imaging (MRI) may reduce visibility of fine anatomical details, motivating computational super-resolution (SR) to enhance perceived image quality. This study proposes an SR pipeline for 2D brain MRI images using a U-Net baseline model and a reinforcement learning (RL) agent to automate hyperparameter tuning. Because the selected public dataset does not provide paired low-resolution/high-resolution (LR–HR) images, LR inputs are generated synthetically using a controlled degradation process (blur–downsample–upsample–noise), with deterministic degradation for validation and testing to ensure stable evaluation. The baseline U-Net is trained using an L1 objective (optionally mixed with differentiable SSIM loss), AdamW optimizer, and ReduceLROnPlateau scheduler guided by validation PSNR. A Double Deep Q-Network (Double DQN) agent then selects discrete action combinations of learning rate and SSIM-weighted loss mixing to fine-tune the baseline. For the held-out test set (n=60), the baseline improves degraded inputs from 27.04±3.21 dB to 30.10±3.59 dB PSNR and from 0.706±0.132 to 0.875±0.064 SSIM, respectively. RL fine-tuning yields a modest additional PSNR gain to 30.20±3.58 dB and SSIM remains comparable at 0.873±0.066. The paired statistical tests confirm that the PSNR improvement is significant ($p<0.01$), while changes in SSIM are not statistically significant, suggesting that for the tested synthetic degradation setting RL can provide reliable but incremental refinement when the baseline is already strong.

Keywords: Medical image super-resolution; Brain MRI; Synthetic degradation; U-Net; Reinforcement learning; Hyperparameter optimization; PSNR; SSIM;

1. INTRODUCTION

Magnetic resonance imaging (MRI) is a favored imaging technique for both clinical and research purposes, as it produces excellent contrast for soft tissue and does not expose patients to ionizing radiation. However, several real-world factors may lead to MRIs being acquired and processed in a way that will decrease the final images' resolution, such as limitations in imaging hardware, time constraints, motion artifacts, and the choice of imaging protocols. Loss of spatial resolution can obscure critical anatomical detail and/or the presence of specific tissues which may in turn affect subsequent processes such as visual assessment, tissue segmentation, or quantitative measurement [1], [2], [3].

Computational super-resolution (SR) is a technique that aims to generate a higher quality image by restoring or hallucinating previously unrecorded high-frequency information that is missing from a lower-quality image. While most traditional SR methods are based on some type of image interpolation or prior information, the increasing use of complex, data-driven deep learning models has signified a shift toward SR methods based on this approach as the dominant technique. Particularly in medical imaging, the use of SR as a post-processing technique which does not alter the imaging modality is very appealing. Early convolutional SR networks, deeper residual SR networks, and adversarial networks are representative of the different SR methods for perception-based enhancement, while SR methods based on encoder-decoder structures and U-Net designs are dominant in the biomedical imaging field as they balance the preservation of spatial detail and the extraction of features at multiple scales [3], [4], [5].

Although SR networks have proven to be successful, their performance is heavily contingent upon training configuration, which includes optimization parameters and the design of the loss function. In real-world applications, these parameters are frequently adjusted manually using trial-and-error, which can be inefficient, and result in less-than-ideal configurations. Reinforcement learning (RL) offers an alternative in which an agent learns to select actions—such as discrete hyperparameter choices—based on the observed state of model performance [6], [7].

This paper focuses on a reproducible SR pipeline using a public 2D brain MRI dataset. Since the dataset is not designed for SR and does not provide paired LR–HR images, we

adopt synthetic degradation, where the original images are treated as pseudo-HR targets and degraded inputs are generated in a controlled manner using blur, downsampling, upsampling, and noise. This approach supports proof-of-concept SR studies when paired acquisitions are unavailable and enables stable benchmarking when the validation and test degradations are made deterministic [8], [9].

Most SR studies emphasize network design or large-scale supervised training on paired LR–HR datasets. In small-to-medium data settings, training configuration becomes crucial and is often tuned manually, while RL-based controllers remain less explored in lightweight SR fine-tuning. Moreover, the chosen public 2D brain MRI dataset does not provide paired LR–HR acquisitions, so this work is framed as a proof-of-concept under deterministic synthetic degradation rather than a clinical paired benchmark. Accordingly, we investigate whether a Double DQN agent can provide measurable improvements over a strong U-Net baseline by selecting discrete combinations of learning rate and loss mixing during fine-tuning, and we quantify its impact using PSNR, SSIM, and paired statistical testing [5], [7], [10], [11]. Our contributions are a deterministic synthetic SR pipeline to reduce evaluation noise, a strong U-Net baseline trained with standard optimization practices, and an RL fine-tuning framework with a reproducible discrete action space and reward design, evaluated with per-image metrics and paired significance tests to demonstrate incremental but consistent benefits in the tested synthetic setting [3], [12], [13], [14].

2. METHODS

This part elaborates on the details of the dataset, the synthetic degradation methodology for creating pairs of inputs and targets, the U-Net architecture and its training methodology which serves as the base, the Double DQN fine-tuning method, as well as the evaluation and statistical method descriptions. Overall, the proposed framework integrates synthetic degradation, supervised U-Net training, and Double DQN fine-tuning for reproducible evaluation. The framework is proposed and evaluated under a deterministic synthetic degradation setting.

The overall workflow of this study proceeds as follows. First, the public 2D brain MRI dataset is collected, converted to grayscale, resized to 256×256, normalized to [0,1], and

split into training, validation, and test sets using stratified sampling. Second, because paired LR–HR acquisitions are unavailable, LR inputs are generated from HR targets using a controlled synthetic degradation process (Gaussian blur, ×2 downsampling, bicubic upsampling, and additive Gaussian noise), with random degradation for training and deterministic degradation for validation and testing.

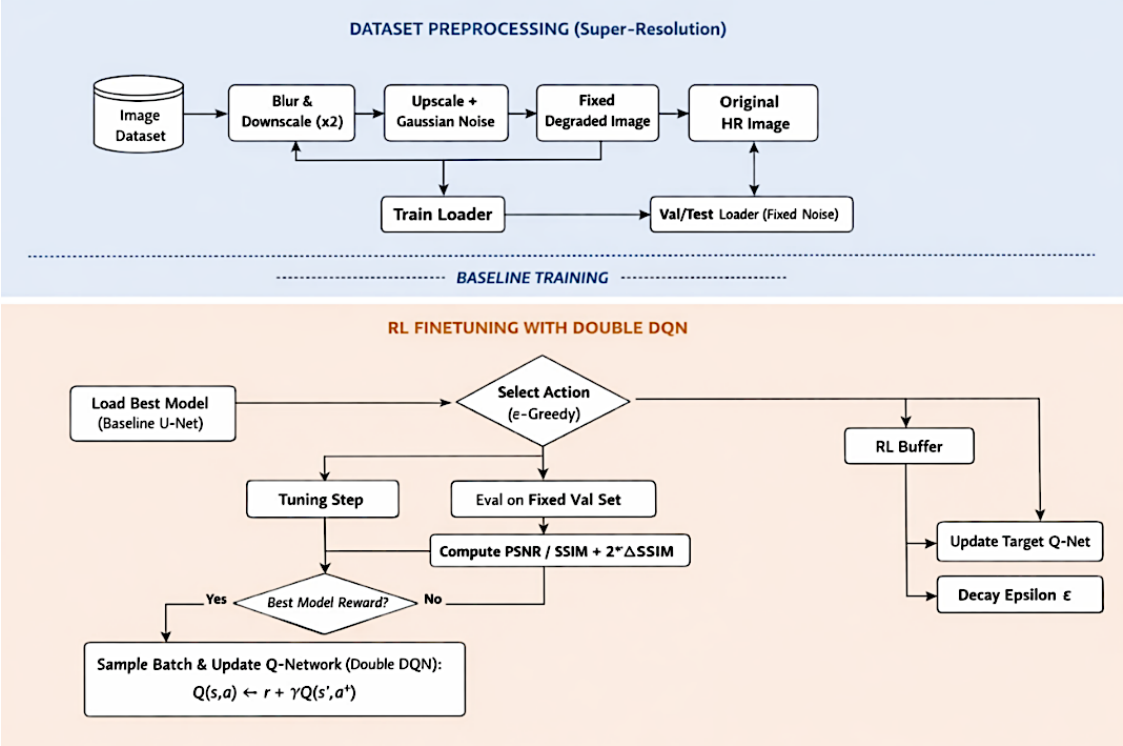

Figure 1. Overall workflow of the proposed RL-guided U-Net super-resolution

Third, a baseline U-Net SR model is trained in a supervised manner to map LR-up inputs to HR targets using L1 loss (optionally mixed with differentiable SSIM), AdamW optimization, and PSNR-driven learning-rate scheduling, and the best checkpoint is selected by validation PSNR. Fourth, the converged baseline is fine-tuned using a Double DQN agent that selects discrete action combinations of learning rate and the SSIM-loss mixing coefficient based on the current PSNR/SSIM state, with rewards defined by improvements relative to the baseline.

Finally, per-image PSNR and SSIM are reported for the benchmark as well as RL-fine-tuned models on the held-out test set, as well as paired statistical tests (paired t-test &

Wilcoxon signed-rank test) to determine if observed differences are statistically significant.

2.1. Dataset and Data Split

We utilized a publicly available dataset which has been organized into 2D brain MRI images, split into two classes (Normal and Tumor). While the dataset has existing labels, the SR task in this study concerns image restoration, and thus the labels are solely for stratified splitting to preserve the class distribution across the train, validation, and test sets. Following the split protocols, the dataset was divided into train/validation/test sets in a 70%/15%/15% ratio, whereby the test set had 60 images. Each image was converted to grayscale, downsampled to 256 × 256 pixels, and normalized to the [0,1] range using the standard tensor conversion [15]. Because this dataset was originally curated for classification rather than paired super-resolution, we repurpose it for SR by generating synthetic LR–HR pairs via controlled degradation; consequently, the results should be interpreted as a proof-of-concept under the specified degradation model.

2.2. Synthetic Degradation Model

Given the absence of paired LR-HR images in the dataset, we formulate a supervised SR case by considering each downscaled source image as an LR input and each original image as an HR target. The LR input is constructed by performing Gaussian blurring, downscaling by a constant factor (x2 as reported in the experiments), upsampling by bicubic interpolation to 256x256, and the addition of a small amount of Gaussian noise to simulate acquisition noise. For each training sample, the blurring and noise to be added are selected randomly. For the validation and test phases, the degradation is performed deterministically by using fixed indices. Thus, the same degraded input is generated for each epoch, which results in more stable estimates for the RL reward [4], [5], [9].

2.3. Baseline Network Architecture (U-Net)

Figure 2 illustrates the U-Net architecture used as the baseline SR model for 2D grayscale images. It takes as input a 1×256×256 LR-up image and passes it through three encoder tiers of DoubleConv blocks, which have progressively larger channel counts, and a MaxPool for downsampling after each. The deepest multi-scale representation is

captured by a bottleneck block. Each level of the decoder upsamples the features to the original resolution, and to maintain spatial resolution, features from the corresponding encoder level are concatenated through skip connections. The output is a 1×256×256 image after the last 1×1 convolution. Skip connections are used to mitigate oversmoothing and enhance the reconstruction of edges [16], [17], [18], [19], [20].

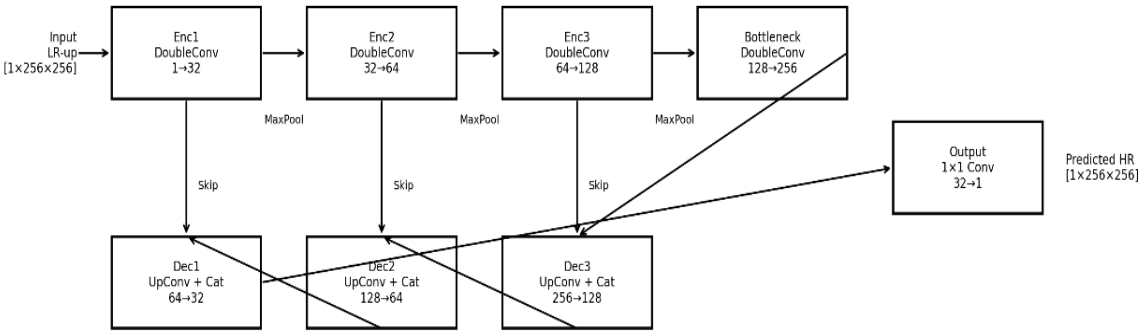

Figure 2. U-Net architecture used as the supervised SR baseline for brain MRI

2.4. Baseline Training Strategy

The baseline model is trained in a supervised way to map the LR-up inputs to HR targets. The primary loss is L1Loss (mean absolute error). When differentiable SSIM is available, we also consider a mixed objective of the form: $L = (1 - \alpha)\cdot L1 + \alpha\cdot(1 - SSIM(\hat{y}, y))$, where $\alpha \in [0,1]$ controls the SSIM contribution. Optimization uses AdamW with an initial learning rate of 1e-4 and weight decay of 1e-4. ReduceLROnPlateau monitors validation PSNR (mode=max) and reduces the learning rate by a factor of 0.5 after several epochs without improvement. The best baseline checkpoint is selected by validation PSNR and used to initialize RL fine-tuning [4], [21], [22], [23].

2.5. Reinforcement Learning Fine-Tuning (Double DQN)

After the baseline training has converged, we proceed to fine-tune the model with the help of a Reinforcement Learning (RL) agent, which chooses some hyperparameters to act on. The agent does not alter the U-Net architecture, rather, it determines which training configuration is to be applied at which step. The action space comprises six combinations of a learning rate (1e-4 and 5e-5) and a loss-mixing coefficient alpha (0.1, 0.3, and 0.5). Restricted options have been empirically shown to improve the stability of

the agent and reduce the probability of it selecting extreme configurations that negatively impact performance [7], [24].

Design rationale: The discrete action space is kept small to help stabilize learning and to limit the search into plausible configurations in a small-data setting. The learning-rate candidates (1e-4 and 5e-5) correspond to the baseline training scale plus a more conservatively set refinement step, while the SSIM-mixing candidates (α = 0.1, 0.3, 0.5) give a low-to-high balance between pixel fidelity and structural similarity.

The state uses PSNR and SSIM because they capture complementary aspects of reconstruction quality (fidelity and structure); PSNR is normalized by 50 so both state components lie on comparable numeric ranges. The reward is defined relative to the baseline on the same data to measure incremental improvement and reduce sensitivity to absolute metric scale; SSIM is up-weighted to encourage structure preservation in addition to PSNR gains [7], [12], [25].

Table 1. The six discrete actions (index, learning rate, alpha)

Action index	Learning rate (lr)	Loss-mixing coefficient (α)
0	1e−4	0.1
1	1e−4	0.3
2	1e−4	0.5
3	5e−5	0.1
4	5e−5	0.3
5	5e−5	0.5

The state is defined as s = [PSNR/50, SSIM], where PSNR is scaled by 50 to keep it on a comparable numeric range to SSIM. The RL reward is defined relative to the baseline model and is computed from the change in PSNR and SSIM between the baseline and the next iteration, i.e., R = (PSNR_next − PSNR_base) + 2 * (SSIM_next − SSIM_base). We utilize Double DQN to learn the action-value function Q(s,a). While the online Q-network determines the next action, a different target network assesses the chosen action to mitigate the overestimation of Q-values. Experience tuples (s, a, r, s_next) are kept in a replay buffer and are randomly sampled to update the Q-network [10], [13], [26].

2.6. Evaluation Metrics

We evaluate image quality using PSNR and SSIM. PSNR summarizes pixel-level fidelity, where higher values indicate lower mean squared error between prediction and target. SSIM summarizes structural similarity, where higher values indicate more similar luminance, contrast, and structure. Metrics are computed per image on the test set and summarized as mean ± standard deviation [4], [9], [22], [27], [28].

2.7. Statistical Testing

Because baseline and RL outputs are produced for the same test images, we use paired statistical tests to assess whether differences are reliable. We report the paired t-test for mean differences and the Wilcoxon signed-rank test as a non-parametric alternative. Statistical significance is reported using p-values [3].

2.8. Implementation Details

All models and training loops are implemented in PyTorch. Images are processed as single-channel grayscale tensors normalized to the [0,1] range. The HR target size is fixed at 256×256. Synthetic SR uses a scale factor of ×2, meaning that images are downsampled to 128×128 and then upsampled back to 256×256 before being fed to the network. Gaussian blur and Gaussian noise are applied as part of the synthetic degradation. In the implementation, blur sigma is sampled within approximately 0.8 to 1.6 and the noise standard deviation is sampled within approximately 0.00 to 0.05. These values define the restoration difficulty and should be explicitly reported because they affect absolute PSNR and SSIM.

The baseline U-Net is trained for 40 epochs using AdamW with an initial learning rate of 1e-4 and weight decay of 1e-4. ReduceLROnPlateau monitors validation PSNR and reduces the learning rate by a factor of 0.5 if PSNR does not improve for several epochs. To improve stability, predicted outputs are clamped to [0,1] during training and evaluation so that PSNR and SSIM are computed consistently with a data range of 1.0. Gradient clipping is applied with a modest norm threshold to reduce the risk of instability from occasional gradient spikes.

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

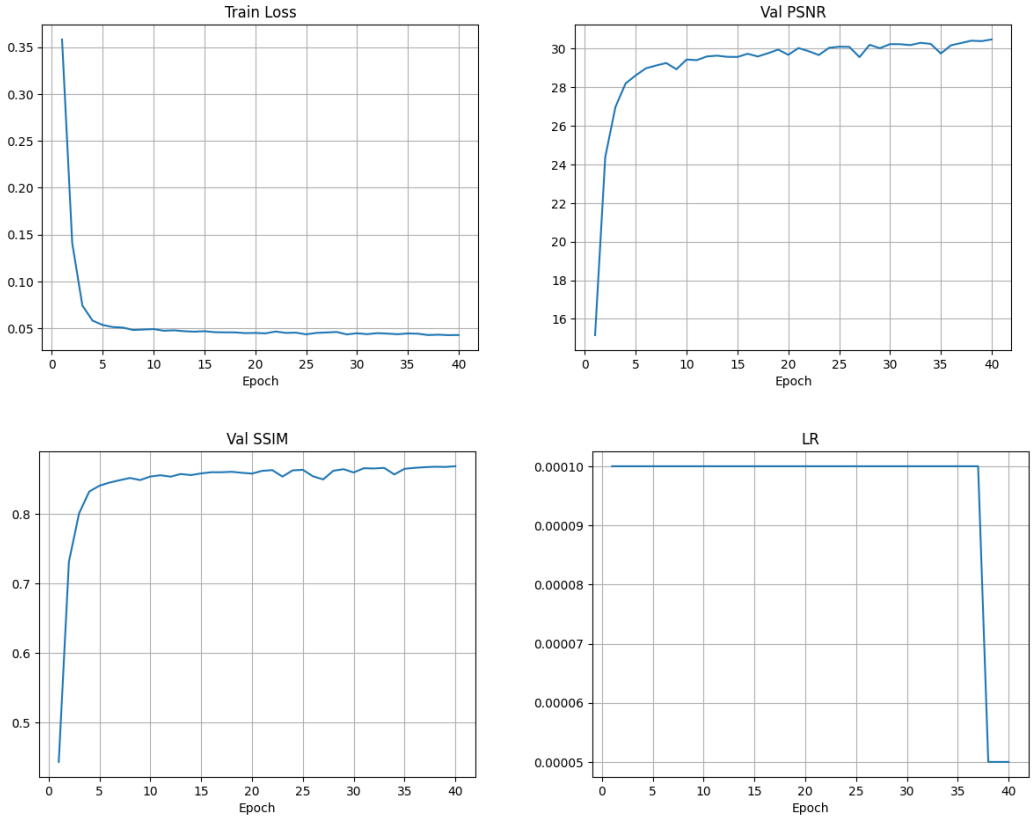

Figure 3. Train loss, validation PSNR, validation SSIM, and learning rate schedule over 40 epochs.

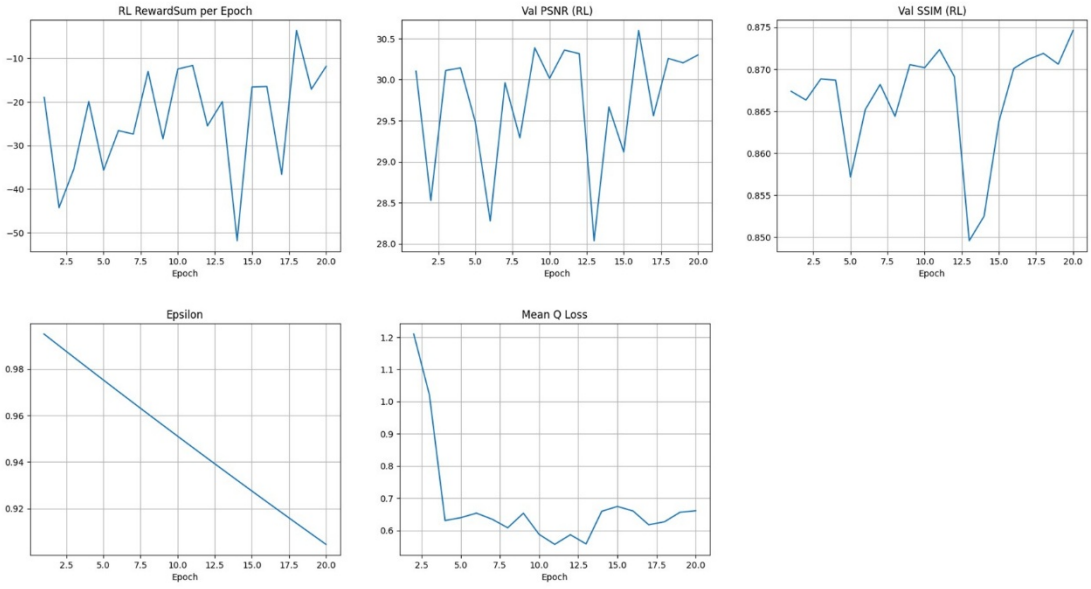

Figure 4. RL RewardSum per epoch, validation PSNR/SSIM during RL fine-tuning, greedy exploration decay, and mean Q-loss across updates.

For RL fine-tuning, the model is initialized from the best baseline checkpoint. The Double DQN controller uses a small fully connected Q-network with two hidden layers of 64 units and ReLU nonlinearities. A replay buffer stores experience tuples and is sampled in mini-batches to update the Q-network. The discount factor is set to 0.99 and the target network is updated by soft updates with tau set to 0.01. Exploration is controlled by an epsilon-greedy strategy in which epsilon starts at 1.0 and decays gradually toward a minimum of 0.05. In the reported configuration, RL fine-tuning is run for 20 epochs, using the same dataset and deterministic validation evaluation [6], [13].

2.9. Baselines for Comparison

We compare three conditions: (i) the degraded input (LR-up) as a no-restoration baseline; (ii) the supervised baseline U-Net; and (iii) RL-UNet (the baseline fine-tuned using Double DQN-selected hyperparameters).

3. RESULTS AND DISCUSSION

3.1. Quantitative Results on the Test Set

Table 2 provides an overview of the mean results (plus standard deviation) for each image in the given dataset (n=60). The results of the degraded inputs show PSNR 27.04 ± 3.21 dB and SSIM 0.706 ± 0.132, representing the quality decrease expected with blur–downsample–upsample–noise degradation. Post supervised training, the baseline U-Net shows an increase in the quality of the images to PSNR 30.10 ± 3.59 dB and SSIM 0.875 ± 0.064. This increase shows that the baseline restores an important amount of the lost high frequency detail and improves the structural coherence, which is consistent with the performance of encoder-decoder architectures and skip connections for restoration tasks.

Table 2. Quantitative reconstruction performance on the test set (n=60).

Method	PSNR (dB) mean ± SD	SSIM mean ± SD	ΔPSNR vs Baseline (dB)	ΔSSIM vs Baseline
LR-up (Degraded input)	27.04 ± 3.21	0.706 ± 0.132	—	—
Baseline U-Net	30.10 ± 3.59	0.875 ± 0.064	—	—
RL-UNet (Double DQN fine-tuned)	30.20 ± 3.58	0.873 ± 0.066	+0.102 ± 0.268	−0.0022 ± 0.0131

The RL-UNet achieves PSNR 30.20 ± 3.58 dB and SSIM 0.873 ± 0.066 on the test set. Relative to the baseline, the average paired change is small (ΔPSNR ≈ +0.102 dB and ΔSSIM ≈ −0.0022), which is expected when the baseline has already converged to a strong solution and the RL controller is restricted to a small, safe action space. Overall, the results indicate that the RL agent can slightly refine pixel-level fidelity without meaningfully changing structural similarity under this configuration [14], [24]. Standard deviations can be interpreted as an indication of how difficult the images in the dataset are. Higher PSNR/SSIM are results of samples that are more easily segmented or provide high contrast, while lower results are of samples that have more difficult to identify edges or complex textures. The 3 dB PSNR variation across methods is evidence of this spread and demonstrates that the improvement is not consistent across all images.

3.2. Distribution of Improvements (ΔPSNR and ΔSSIM)

As shown in Figure 5, the ΔPSNR histogram exhibits a mild shift toward positive values with many samples clustered near zero, indicating small improvements distributed across a large portion of the test set. In contrast, the ΔSSIM histogram is tightly centered near zero with a slightly negative mean, suggesting that structural similarity is largely unchanged under the tested configuration and that observed gains are primarily reflected in pixel-level fidelity (PSNR).

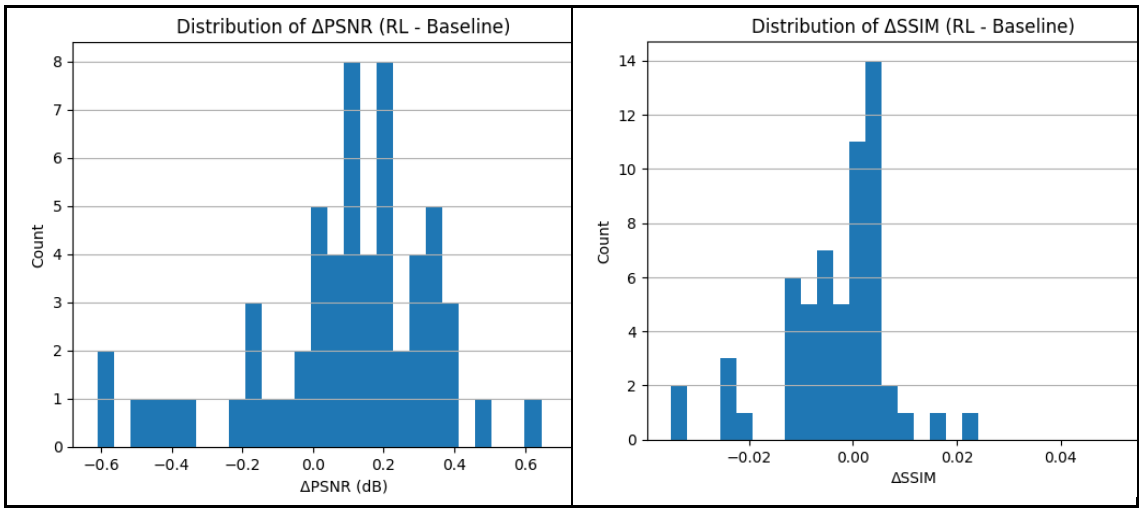

Figure 5. Distributions of per-image paired differences between RL-UNet and the baseline U-Net across the test set (n=60): (a) ΔPSNR and (b) ΔSSIM.

Negative differences can still occur on individual images. Fine-tuning may slightly over-specialize to characteristics of the synthetic degradation or introduce minor trade-offs between pixel-level and structural measures for specific samples. However, the overall distributions indicate that most per-image changes are small in magnitude, with only a limited number of larger outliers.

3.3. Statistical Significance of RL Improvements

Although the improvements from RL are practically incremental, paired statistical tests show that the PSNR gain is statistically reliable. Across the 60 test images, the paired differences are ΔPSNR 0.102 ± 0.268 dB and ΔSSIM −0.0022 ± 0.0131. For PSNR, the paired t-test yields p=0.004572 and the Wilcoxon signed-rank test yields p=0.001199, indicating a consistent improvement. For SSIM, neither test indicates a significant difference (paired t-test p=0.1997; Wilcoxon p=0.1239). These results support the interpretation that RL provides a small but dependable PSNR refinement when the baseline is already strong, while SSIM remains statistically unchanged.

Table 3. Paired statistical tests for RL improvements relative to the baseline U-Net (n=60). Mean paired differences are reported as mean ± standard deviation.

Metric	Mean Δ (RL−Baseline) ± SD	Paired t-test p-value	Wilcoxon signed-rank p-value
PSNR (dB)	+0.102 ± 0.268	0.004572	0.001199
SSIM	−0.0022 ± 0.0131	0.1997	0.1239

This conclusion is consistent with the histogram analysis: when small changes occur consistently across many paired samples, they can yield statistically significant results, as observed for PSNR. No stronger interpretation is warranted; under the present setup, RL mainly provides a dependable and automated mechanism to obtain small PSNR refinements when a strong supervised baseline is already available. Table 3 reports p-values alongside mean ΔPSNR and mean ΔSSIM. From a practical medical-image perspective, a ~0.1 dB PSNR gain without a significant SSIM gain suggests that the RL stage functions mainly as automated, conservative hyperparameter refinement rather than a qualitatively transformative enhancement of perceived structure.

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

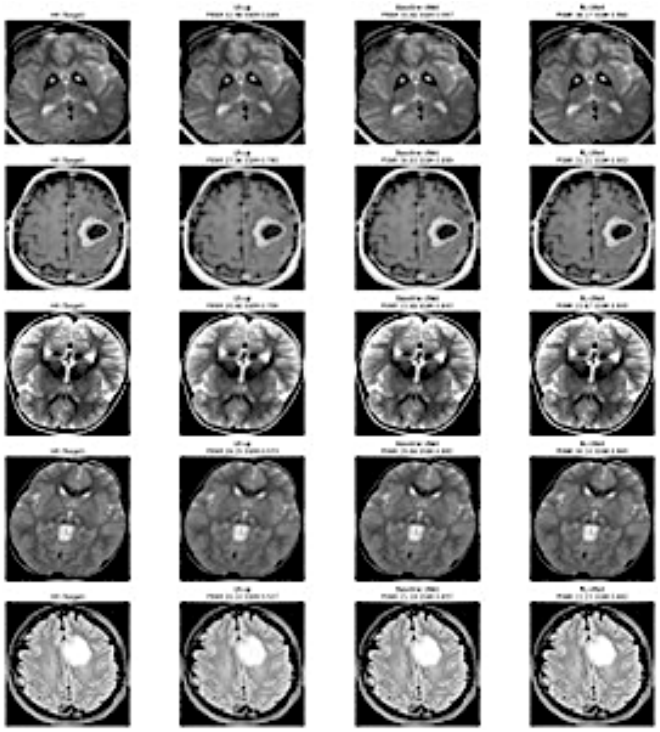

Figure 6. Qualitative comparison on representative test slices.

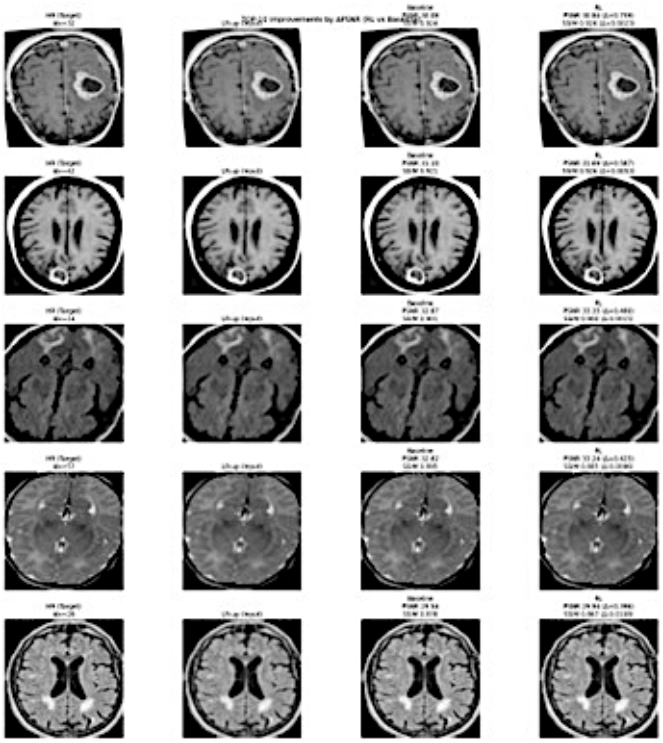

Figure 7. Top-5 test images by metric improvement after RL fine-tuning, shown as HR target, LR-up input, baseline output, and RL output: (a) ranked by ΔPSNR and (b) ranked by ΔSSIM (RL−Baseline).

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

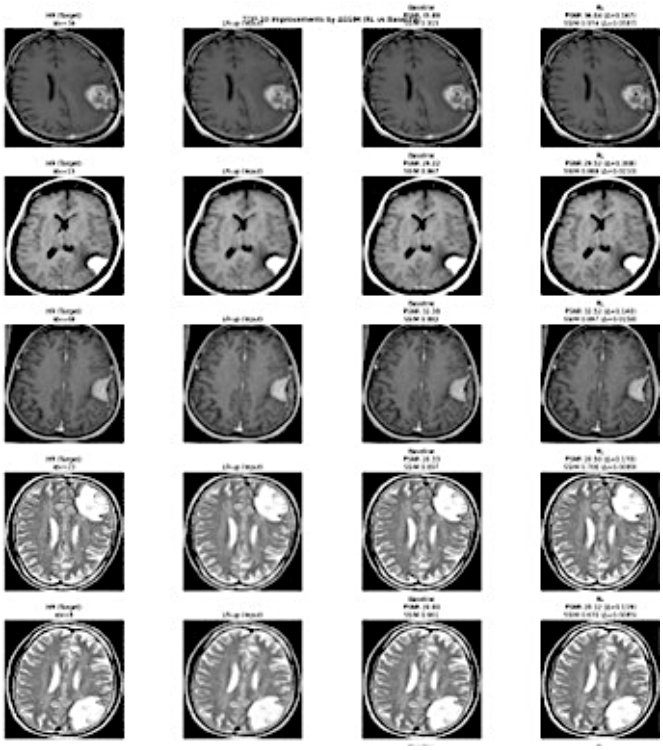

Figure 8. Bottom-5 (worst) test images by metric change after RL fine-tuning: (a) ranked by ΔPSNR and (b) ranked by ΔSSIM (RL−Baseline).

3.4. Discussion

The results indicate that the primary reconstruction gain in this study comes from the supervised U-Net baseline, while the Double DQN fine-tuning stage contributes a smaller but statistically reliable refinement in pixel-level fidelity rather than a broad structural improvement. As shown in Table 2, the degraded input starts at 27.04 ± 3.21 dB PSNR and 0.706 ± 0.132 SSIM, whereas the baseline U-Net improves performance substantially to 30.10 ± 3.59 dB and 0.875 ± 0.064, confirming that the encoder-decoder architecture restores much of the lost high-frequency detail and structural coherence. The RL-UNet then increases mean PSNR slightly further to 30.20 ± 3.58 dB, but SSIM changes marginally to 0.873 ± 0.066. This interpretation is reinforced by the paired statistical analysis in Table 3, where the PSNR gain is significant ($p = 0.004572$ by paired t-test; $p = 0.001199$ by Wilcoxon), while the SSIM difference is not statistically significant. Taken together, these findings suggest that under the present synthetic degradation setting, the RL stage should be interpreted as a conservative optimization layer rather than a qualitatively transformative reconstruction method [14], [24].

The distribution plots in Figure 5 further clarify the nature of this gain. The ΔPSNR histogram shows a mild shift toward positive values, with many samples clustered close to zero, indicating that improvements are generally small but widespread rather than dramatic and isolated. By contrast, the ΔSSIM histogram is tightly centered around zero with a slightly negative mean, which shows that the RL stage does not systematically improve structural similarity across the test set. This pattern is important because it explains why the PSNR improvement can be statistically significant even though the visual changes are subtle: when many images experience a small positive shift, the average effect can still be dependable. At the same time, the narrow spread around zero also shows that the RL policy does not destabilize the model or produce erratic behavior across the dataset. The overall effect is therefore best understood as incremental refinement, not a second-stage reconstruction that changes the qualitative character of the output.

The qualitative examples in Figure 6 are consistent with this reading of the quantitative results. The degraded inputs show the expected effects of the synthetic pipeline, including blurred anatomical boundaries, softened edges, and reduced local contrast. The baseline U-Net restores most of the visible improvement, producing sharper contours and more coherent tissue structures. Relative to this baseline, the RL-UNet output is usually visually very similar, with only localized differences such as slightly reduced residual blur, mild sharpening of fine boundaries, or small contrast adjustments in selected regions. In other words, almost all of the obvious visual recovery comes from the supervised model, while the RL stage acts mainly as a fine adjustment mechanism. This is fully consistent with the statistical outcome: a small average gain in PSNR, paired with no significant improvement in SSIM, implies that the RL stage refines pixel fidelity in ways that are measurable but not consistently strong enough to alter perceived structural similarity. Thus, the qualitative comparison should be interpreted as confirming, rather than contradicting, the statistical findings.

The Top-5 cases in Figure 7 provide additional insight into where RL appears to help most. These cases often involve slices in which degradation has weakened thin boundaries or where the baseline remains slightly oversmoothed in localized regions. In such examples, RL fine-tuning appears to recover minor edge sharpness or improve local contrast just enough to yield measurable gains in ΔPSNR, and in some cases, modest

improvements in ΔSSIM. This suggests that the RL controller is most beneficial when the baseline has already produced a strong reconstruction but still leaves small correctable errors in fine detail representation. The pattern also aligns with the restricted action space used in this study. Because the RL agent is allowed only a small, safe set of hyperparameter adjustments, it is unlikely to produce radical changes; instead, it nudges the optimization process toward slightly better local solutions. This helps explain why the improvements are real yet limited in scale. In practical terms, the RL component seems most useful as an automated fine-tuning policy that extracts marginal additional quality from a strong baseline without requiring repeated manual hyperparameter search.

The Bottom-5 cases in Figure 8 highlight the boundaries of this benefit and help explain why the SSIM gains were not significant. In these underperforming samples, RL fine-tuning can introduce subtle over-sharpening, minor texture amplification, or small structural trade-offs that slightly reduce similarity to the ground truth. These failure modes are not catastrophic, but they show that the RL policy does not help every image equally. Negative paired differences are therefore not surprising, especially in a setting where the baseline is already strong and the remaining room for improvement is small. In such conditions, some images may be more sensitive to slight shifts in learning rate or loss-mixing behavior, causing the policy to over-specialize to characteristics of the synthetic degradation rather than improve general structural fidelity. This observation matches the histogram behavior in Figure 5, where most changes are near zero and only a limited number of images show larger positive or negative deviations. It also explains why the RL stage can improve PSNR on average while leaving SSIM statistically unchanged: the policy is refining certain intensity-level details, but not doing so uniformly enough to produce a reliable gain in global structural similarity.

In relation to the research objectives, the findings show that the proposed RL framework succeeds in its intended role, but that role is narrower than that of the supervised reconstruction network. The objective was not to replace the baseline U-Net with reinforcement learning, but to test whether Double DQN-based hyperparameter control could provide additional performance gains once a strong supervised solution had already been established. Under that interpretation, the results are encouraging. The RL stage delivered a small, consistent, and statistically significant PSNR improvement, which

indicates that the agent was able to learn useful fine-tuning behavior from the reward signal. At the same time, the absence of a significant SSIM improvement shows that the practical impact of that learned behavior is limited mainly to pixel-level fidelity. This distinction is important. It suggests that RL is most valuable here as an automation mechanism for conservative optimization, reducing the burden of manual tuning rather than serving as a standalone source of major reconstruction improvement. The constrained discrete action space also appears to have contributed positively by preventing unstable or excessively large hyperparameter changes, thereby keeping the fine-tuning process safe and controlled.

From a practical standpoint, the study implies that the largest performance gains in small-data super-resolution pipelines still come from building a strong supervised baseline, including the choice of architecture, loss formulation, optimizer configuration, and learning-rate strategy. Once that baseline is near convergence, RL can provide additional benefit, but the magnitude of that benefit is likely to be modest. In this sense, RL should be viewed as an incremental refinement tool that can automate part of the tuning process and recover small improvements that might otherwise require repeated trial-and-error experimentation. The use of deterministic degradation for validation and testing is also justified in this context, because it stabilizes PSNR/SSIM measurement and makes it easier to detect subtle gains. This is especially relevant when the expected improvement is on the order of 0.1 dB, where noisy evaluation conditions could easily obscure the signal. The current findings therefore support a workflow in which supervised learning delivers the dominant restoration effect, while RL is layered on top as a controlled and efficient fine-tuning strategy.

Several limitations should temper the interpretation of these results. The LR-HR pairs are synthetically generated, so the findings primarily reflect the defined blur-downsample-upsample-noise process rather than naturally paired clinical acquisitions. The dataset is also small and 2D, which limits generalizability and leaves open the question of whether the same refinement pattern would hold in larger, more heterogeneous, or volumetric medical imaging settings. In addition, the RL agent operates over a restricted action space, selecting only the learning rate and loss-mixing coefficient. While this was useful for stability, it also limits the scope of what the agent can optimize. A richer action space, more task-aware reward formulation, or evaluation against downstream clinical tasks

could potentially reveal stronger or more meaningful benefits. Accordingly, future work should validate the approach on larger datasets, realistic degradations, and possibly 3D reconstruction settings, while also exploring broader RL control policies and task-level metrics beyond PSNR and SSIM.

4. CONCLUSION

Synthetic degradation is used in this study to construct paired input–target images and enable a reproducible super-resolution pipeline for 2D brain MRI under a deterministic synthetic evaluation setting. A strong supervised U-Net baseline trained with L1 loss (optionally mixed with differentiable SSIM), AdamW optimization, and PSNR-driven learning-rate scheduling provides the main reconstruction gains, improving PSNR from 27.04 ± 3.21 dB to 30.10 ± 3.59 dB and SSIM from 0.706 ± 0.132 to 0.875 ± 0.064 on the held-out test set. Building on this baseline, Double DQN fine-tuning is used as a constrained hyperparameter controller that selects from a finite set of learning rates and loss-mixing coefficients, yielding a modest PSNR refinement to 30.20 ± 3.58 dB (ΔPSNR 0.102 ± 0.268 dB) while SSIM remains comparable at 0.873 ± 0.066 (ΔSSIM −0.0022 ± 0.0131); paired tests indicate the PSNR gain is statistically significant whereas SSIM changes are not. Overall, these results support the conclusion that RL can reduce manual trial-and-error by providing a stable, automated fine-tuning mechanism that delivers small but reliable PSNR improvements when the supervised baseline is already strong, while broader clinical generalizability will require evaluation under more realistic MRI degradation conditions and, when available, paired clinical LR–HR acquisitions and larger datasets, as well as downstream task-based assessments (e.g., segmentation or classification) with clinically aligned reward definitions.

REFERENCES

[1] T. C. Arnold, C. W. Freeman, B. Litt, and J. M. Stein, "Low-field MRI: Clinical promise and challenges," *J. Magn. Reson. Imaging*, vol. 57, no. 1, pp. 25–44, Jan. 2023, doi: 10.1002/jmri.28408.

[2] M. W. Haskell, J. F. Nielsen, and D. C. Noll, "Off-resonance artifact correction for MRI: A review," *NMR Biomed.*, vol. 36, no. 5, May 2023, doi: 10.1002/nbm.4867.

[3] C. Sarasaen, S. Chatterjee, M. Breitkopf, G. Rose, A. Nürnberger, and O. Speck, "Fine-tuning deep learning model parameters for improved super-resolution of dynamic MRI with prior-knowledge," *Artif. Intell. Med.*, vol. 121, Nov. 2021, doi: 10.1016/j.artmed.2021.102196.

[4] K. Chauhan *et al.*, "Deep Learning-Based Single-Image Super-Resolution: A Comprehensive Review," *IEEE Access*, vol. 11, pp. 21811–21830, 2023, doi: 10.1109/ACCESS.2023.3251396.

[5] H. Su *et al.*, "A review of deep-learning-based super-resolution: From methods to applications," *Pattern Recognit.*, vol. 157, Jan. 2025, doi: 10.1016/j.patcog.2024.110935.

[6] D. Passos and P. Mishra, "A tutorial on automatic hyperparameter tuning of deep spectral modelling for regression and classification tasks," *Chemom. Intell. Lab. Syst.*, vol. 223, Apr. 2022, doi: 10.1016/j.chemolab.2022.104520.

[7] F. M. Talaat and S. A. Gamel, "RL based hyper-parameters optimization algorithm (ROA) for convolutional neural network," *J. Ambient Intell. Humaniz. Comput.*, vol. 14, no. 10, pp. 13349–13359, Oct. 2023, doi: 10.1007/s12652-022-03788-y.

[8] P. Nandal, S. Pahal, A. Khanna, and P. Rogerio Pinheiro, "Super-Resolution of Medical Images Using Real ESRGAN," *IEEE Access*, vol. 12, pp. 176155–176170, 2024, doi: 10.1109/ACCESS.2024.3497002.

[9] J. Y. Lee, M. I. Hussain, K. H. Lee, H. S. Shim, S. H. Han, and D. Yang, "Transfer Learning-Based Super-Resolution for High-Precision Medical Imaging," *IEEE Access*, vol. 13, pp. 124776–124791, 2025, doi: 10.1109/ACCESS.2025.3587263.

[10] Y. Yu, Y. Liu, J. Wang, N. Noguchi, and Y. He, "Obstacle avoidance method based on double DQN for agricultural robots," *Comput. Electron. Agric.*, vol. 204, Art. no. 107546, 2023, doi: 10.1016/j.compag.2022.107546.

[11] A. Ullah, I. Ullah, Q. M. ul Haq, S. Rubab, J. Baili, and M. A. Khan, "SDN-driven multi-objective task offloading in IoT-enabled UAVs in edge-cloud computing using double DQN," *IEEE Trans. Consum. Electron.*, 2025.

[12] Y. Chen, R. Xia, K. Yang, and K. Zou, "MICU: Image super-resolution via multi-level information compensation and U-net," *Expert Syst. Appl.*, vol. 245, p. 123111, Jul. 2024, doi: 10.1016/j.eswa.2023.123111.

[13] A. Ly, R. Dazeley, P. Vamplew, F. Cruz, and S. Aryal, "Elastic step DQN: A novel multi-step algorithm to alleviate overestimation in Deep Q-Networks," *Neurocomputing*, vol. 576, Apr. 2024, doi: 10.1016/j.neucom.2023.127170.

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

[14] B. Luo, Z. Wu, F. Zhou, and B.-C. Wang, "Human-in-the-loop reinforcement learning in continuous-action space," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 35, no. 11, pp. 15735–15744, 2023.

[15] Antor Mahamudul Hashan, "Brain MRI Images." Kaggle. doi: 10.34740/KAGGLE/DS/1250260.

[16] S. Ye, S. Zhao, Y. Hu, and C. Xie, "Single-Image Super-Resolution Challenges: A Brief Review," *Electron. Switz.*, vol. 12, no. 13, Jul. 2023, doi: 10.3390/electronics12132975.

[17] B. Hemanth Sai, S. Mukherjee, and S. R. Dubey, "Adaptive adam-based optimizers using second-order weight decoupling and gradient-aware weight decay for vision transformer," *Mach. Vis. Appl.*, vol. 36, no. 3, p. 68, May 2025, doi: 10.1007/s00138-025-01686-9.

[18] Y. Chen *et al.*, "Deep-learning–based optical coherence tomography reconstruction for high-speed and contrast morphology and vasculature imaging," *J. Biomed. Opt.*, vol. 31, no. 2, Art. no. 025001, 2026, doi: 10.1117/1.JBO.31.2.025001.

[19] N. Siddique, S. Paheding, C. P. Elkin, and V. Devabhaktuni, "U-net and its variants for medical image segmentation: A review of theory and applications," *IEEE Access*, vol. 9, pp. 82031–82057, 2021.

[20] I. Boucherit and H. Kheddar, "Reinforced Residual Encoder–Decoder Network for Image Denoising via Deeper Encoding and Balanced Skip Connections," *Big Data Cogn. Comput.*, vol. 9, no. 4, Apr. 2025, doi: 10.3390/bdcc9040082.

[21] S. H. Park, Y. S. Moon, and N. I. Cho, "Perception-oriented single image super-resolution using optimal objective estimation," in Proc. IEEE/CVF Conf Computer Vision and Pattern Recognition (CVPR), 2023, pp. 1725–1735. doi: 10.1109/CVPR52729.2023.00172.

[22] L. Lin, H. Chen, E. E. Kuruoglu, and W. Zhou, "Robust structural similarity index measure for images with non-Gaussian distortions," *Pattern Recognit. Lett.*, vol. 163, pp. 10–16, 2022.

[23] Q. Ning, W. Dong, X. Li, J. Wu, and G. Shi, "Uncertainty-driven loss for single image super-resolution," *Adv. Neural Inf. Process. Syst.*, vol. 34, pp. 16398–16409, 2021.

[24] J. Feng, Y. Shi, G. Qu, S. H. Low, A. Anandkumar, and A. Wierman, "Stability constrained reinforcement learning for decentralized real-time voltage control," *IEEE Trans. Control Netw. Syst.*, vol. 11, no. 3, pp. 1370–1381, 2023.

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

[25] Y. Y. Tsai, B. Xiao, E. Johns, and G. Z. Yang, "Constrained-Space Optimization and Reinforcement Learning for Complex Tasks," *IEEE Robot. Autom. Lett.*, vol. 5, no. 2, pp. 682–689, Apr. 2020, doi: 10.1109/LRA.2020.2965392.

[26] R. Yu *et al.*, "Improved double DQN with deep reinforcement learning for UAV indoor autonomous obstacle avoidance," *Sci. Rep.*, vol. 15, no. 1, p. 28133, 2025.

[27] C. Chen, Y. Wang, N. Zhang, Y. Zhang, and Z. Zhao, "A Review of Hyperspectral Image Super-Resolution Based on Deep Learning," *Remote Sens.*, vol. 15, no. 11, Jun. 2023, doi: 10.3390/rs15112853.

[28] L. Zhu, B. Zhong, and K.-K. Ma, "APSNR: Artifact Peak Signal-to-Noise Ratio for Image Quality Assessment," *IEEE Trans. Image Process.*, vol. 34, pp. 7180–7192, 2025.