SCHOOL OF ELECTRONICS, ELECTRICAL ENGINEERING and COMPUTER SCIENCE

ECS8056 – THEMED RESEARCH PROJECT

A signed and completed cover sheet must accompany the submission of the Themed Research Project dissertation submitted for assessment.

Work submitted without a cover sheet will **NOT** be marked.

Declaration of Academic Integrity

Before signing the declaration below please check that the submission:

1. Has a full bibliography attached laid out according to the guidelines specified in the Student Project Handbook
2. Contains full acknowledgement of all secondary sources used (paper-based and electronic)
3. Does not exceed the specified page limit
4. Is clearly presented and proof-read
5. Is submitted on, or before, the specified or agreed due date. Late submissions will only be accepted in exceptional circumstances or where a deferment has been granted in advance.
6. Software and files are submitted via Canvas.

I certify that that the submission is my own work, all sources are correctly attributed, and the contribution of any AI technologies is fully acknowledged.

I declare that I have read both the University and the School of Electronics, Electrical Engineering and Computer Science guidelines on plagiarism - https://www.qub.ac.uk/directorates/sgc/learning/LearningResources/Plagiarism/- and that the attached submission is my own original work. No part of it has been submitted for any other assignment and I have acknowledged in my notes and bibliography all written and electronic sources used.

I am aware of the disciplinary consequences of failing to abide and follow the School and Queen's University Regulations on Plagiarism.

Name: (BLOCK CAPITALS) MUHAMMAD SHAMS UL HAQ GHANI

Student Number: 40458280

Student's signature Shams

Date of submission 5/9/2025

“Deep FOCUS: Deep-learning based Focal-loss Optimized Convolution for Ultra-resolution in SMLM”

A dissertation submitted in partial fulfilment of
The requirement for the degree of
MASTER OF SCIENCE in Artificial Intelligence
In
The Queen’s University of Belfast
By

M. Shams Ghani

5th of September, 2025

Deep FOCUS: Deep-learning based Focal-loss Optimized Convolution for Ultra-resolution in SMLM

M Shams Ul Haq Ghani
School of Electronics, EE, and CS
Queen's Univeristy Belfast
Belfast, Northern Ireland, UK
Mghani01@qub.ac.uk

***Abstract*- The fundamental resolution of light microscopes is limited by diffraction. Single Molecule Localization Microscopy (SMLM) overcomes this barrier using techniques like STORM or DNA-PAINT** [4], **which rely on computationally expensive deconvolution-based localization of emitters from diffraction-limited images. Deep learning methods accelerate reconstructions and reduce computational cost, but often compromise accuracy, generalization, and introduce artifacts.**
This study investigates how different loss functions affect deep learning-based SMLM reconstruction. A U-Net architecture was trained on synthetic microscopy data using four different loss functions to evaluate their performance in emitter localization. The results demonstrate that loss function selection critically impacts reconstruction quality. While some loss functions failed to converge or retained imaging artifacts, Focal Loss consistently produced sharp, accurate reconstructions with proper emitter localization. These findings establish the importance of appropriate loss function design for robust, high-resolution deep learning microscopy applications.

I. INTRODUCTION

Light microscopy is one of the most widely used tools in biological research; however, its resolving power is fundamentally constrained by the diffraction limit. This barrier prevents conventional microscopes from resolving structures smaller than ~200nm, limiting the study of subcellular processes [6]. Single-molecule localization microscopy (SMLM) was developed to overcome these limitations, enabling nanoscale imaging beyond the diffraction limit and revolutionizing our understanding of biological systems. SMLM achieves super-resolution through temporal separation: only a sparse subset of the total fluorophore population is emitting at any given time (blinking), creating temporal separation of the fluorescence coming from crowded particles [7].

The technique works by switching on only a small fraction (<1%) of molecules located in a field of view [6] at each moment. Over thousands of frames, different molecules blink on and off, and computational algorithms precisely localize each diffraction-limited spot to nanometre accuracy. The localizations are used to generate a super-resolution image [8] by combining temporal information into a high-resolution reconstruction, enabling visualization of cellular structures at the molecular scale.

Despite its benefits, SMLM remains computationally expensive. Traditional reconstruction approaches like PSF-based Deconvolution, Maximum Likelihood-estimation Method (MLM), or Gaussian fitting are computationally intensive, especially for large datasets or dense emitter regions. The problem also becomes ill-conditioned at higher emitter densities, potentially lacking unique solutions. Deep learning has emerged as a powerful alternative, accelerating reconstructions while reducing computational cost.

However, several challenges remain. Deep learning solutions predict the most likely emitter distribution but do not solve the uniqueness problem. Noise robustness is inconsistent, with high noise leading to false positives. Generalization issues persist; models trained on synthetic data typically fail on experimental data and vice versa. These shortcomings limit practical adoption of deep learning in SMLM workflows.

This study addresses the question: Can we improve Deep Learning performance in SMLM emitter localization through optimization, particularly Loss Function Selection? Through systematic comparative analysis of different loss functions and structural modifications, we seek to identify methods that maximize accuracy, noise robustness, generalization, and efficiency.

II. LITERATURE REVIEW

Recent deep learning approaches for SMLM have advanced the field by significantly improving reconstruction speed and accuracy compared to traditional methods, yet they introduce new challenges in generalization, optimization, and uncertainty handling. Across the literature, a few key points emerge. First, CNN-based architectures, particularly U-Nets, have become the dominant framework for SMLM reconstruction. Deep-STORM demonstrated that a single-stage U-Net trained on simulated data can achieve state-of-the-art resolution under high emitter densities and low signal-to-noise conditions while leveraging GPU acceleration for extreme computational efficiency [1]. Similarly, DECODE extended this paradigm with dual stacked U-Nets incorporating temporal context, improving performance on high-density datasets and providing uncertainty estimates [2]. While both methods achieve high accuracy, DECODE's multi-output architecture introduces greater computational overhead and longer training times compared to Deep-STORM, highlighting a trade-off between model complexity and efficiency.

Second, probabilistic modelling has emerged as a complementary approach for quantifying localization uncertainty. The Bayesian CNN framework by Speiser et al. explicitly models detection and localization probabilistically, providing a principled way to capture uncertainty but at the cost of additional computational complexity [3]. In contrast, DECODE offers practical uncertainty estimates without a full Bayesian formulation, suggesting that deep learning methods can balance probabilistic rigor against training and inference efficiency [2,3].

Third, the generation of reliable training data remains a critical consideration. DNA-PAINT provides a controllable and predictable emitter pattern, enabling high-precision ground truth for network training [4]. The use of such well-characterized datasets underpins the success of methods like Deep-STORM and DECODE, while also highlighting limitations when imaging speed is constrained by binding kinetics. ThunderSTORM serves as a computational baseline, offering multiple traditional localization algorithms, yet remains slow for large, high-density datasets, emphasizing the need for accelerated deep learning solutions [5].

Across these works, one consistent insight is that architectural innovation alone does not guarantee optimal reconstruction. While Deep-STORM and DECODE focus on U-Net designs, and Bayesian CNNs introduce uncertainty modelling, all provide limited systematic exploration of loss function selection [1–3]. This gap is critical: loss functions influence convergence, artifact suppression, and final reconstruction quality. By explicitly investigating four loss functions (L1+L2, MSE, BCE, Focal Loss) in a CNN-based U-Net, the current study demonstrates that focal loss consistently produces sharp, artifact-free reconstructions, whereas traditional losses often fail to converge.

Identified Research Gap: Existing deep learning approaches demonstrate superior speed and accuracy but lack systematic analysis of how loss function selection affects reconstruction quality, convergence behaviour, and artifact generation. Current methods focus on architectural innovations, yet these architectures are often built on standard practices and may not be optimal for SMLM reconstruction, providing limited investigation of optimization fundamentals.

Research Contribution

1. Evaluation of Multiple Loss Functions:

   - Investigated L1+L2, MSE, BCE, and Focal Loss to understand how loss function selection affects reconstruction quality, convergence behaviour, and artifact suppression.

2. **Rationale for Loss Function Selection:**

   - **L1+L2 and MSE:** Commonly used in regression tasks, providing baseline performance for pixel-wise reconstruction.
   - **BCE (Binary Cross-Entropy):** Captures probabilistic differences and is often used in segmentation/localization tasks.
   - **Focal Loss:** Designed to handle class imbalance and emphasize hard-to-predict emitters, making it suitable for sparse, high-density SMLM data.

3. **Key Findings:**

   - Traditional losses (L1+L2, MSE) often fail to converge or produce blurred reconstructions.
   - BCE improves convergence but may introduce minor artifacts.
   - Focal Loss consistently produces sharp, artifact-free reconstructions, demonstrating its critical role in optimizing SMLM deep learning models.

III. METHODOLOGY

3.1 Hardware and Software Configuration

Hardware Platform: All experiments were conducted on a Gigabyte Aorus 16X laptop equipped with an Intel i7-14650HX processor, NVIDIA RTX 4070 GPU, and 16GB RAM. This consumer-grade configuration demonstrates the accessibility of deep learning SMLM approaches without requiring specialized high-performance computing infrastructure.

Software Implementation: Training employed the PyTorch deep learning framework with CUDA acceleration for GPU optimization. [5] ThunderSTORM was utilized for synthetic data generation and baseline performance comparisons, providing standardized evaluation protocols established in the SMLM community.

3.2 Data Generation and Preprocessing

[5] Training data was generated using ThunderSTORM with controlled parameters ensuring reproducibility. Simulation parameters: 128×128-pixel images, 1,500 frames, FWHM 200-350 nm, fluorophore intensity 700-900 photons, emitter density 0.5 μm^{-2} (Fig. 1.). Sample extraction: 8× upsampling, 80 nm camera pixels, 208×208-pixel patches, 500 random patches per frame yielding 10,000 total training examples with minimum 7 emitters per patch. This configuration provides comprehensive training data with known ground truth positions for systematic loss function evaluation.

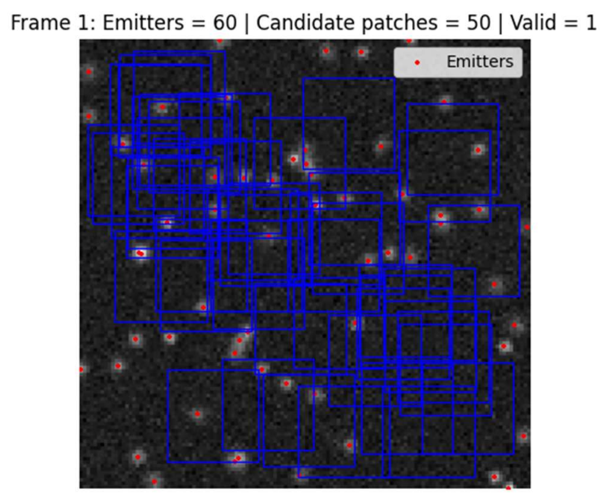

Fig. 1. Candidate patches on a frame

3.3 Initial Model Evaluation

Baseline Assessment: Three pre-trained models were evaluated to establish performance benchmarks: (1) a synthetic data model from the original Deep-STORM implementation [1], (2) an experimental microtubule model, and (3) an external model from the Zenodo repository 6966132. All models were tested on identical input images with corresponding ground truth for consistent comparison (Fig. 2.). Baseline evaluation revealed partial localization capabilities but identified significant limitations, including incomplete emitter detection, spatial imprecision, and elevated noise levels across all pre-trained approaches.

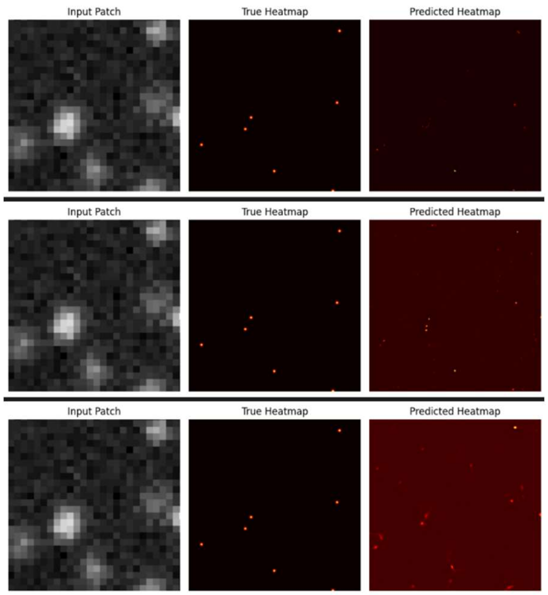

Fig 2. Baseline Model Assessment.
Synthetic (top), Experimental(middle), and Accelerated (bottom)

3.4 Systematic Loss Function Investigation

3.4.1 Original Architecture Training

[1] Four loss functions were systematically evaluated using the original Deep-STORM U-Net architecture: Combined L1+L2, Mean Squared Error (MSE), Binary Cross Entropy (BCE), and Focal Loss. All models were trained on micro-subsets (100 samples) to overfitting, enabling isolation of loss function effects. Each model was evaluated on identical test images with ground truth comparison to ensure consistent assessment conditions.

3.4.2 Modified Architecture Training

The same four loss functions (L1+L2, MSE, BCE, Focal Loss) were re-evaluated using a modified output layer architecture to systematically assess the interplay between architectural design and loss function selection. All models were trained under identical conditions using the same microsubset, enabling direct performance comparison.

The CNN employs a U-Net-inspired encoder-decoder architecture, processing diffraction-limited microscopy patches (208×208 pixels) to produce super-resolved localization maps at 8× upsampling.

Encoder Path: Four encoding stages progressively increase channel dimensions (1→32→64→128→512) using 3×3 convolutions with orthogonal weight initialization, batch normalization, and ReLU activation. Each stage includes 2×2 max pooling for spatial downsampling, reducing spatial dimensions by factors of 2, 4, and 8.

Decoder Path: Three decoding stages use nearest-neighbour upsampling (scale factor 2) followed by convolution blocks, progressively reducing channels (512→128→64→32). The decoder restores the original input dimensions while reconstructing spatial resolution.

Output Configuration: Two variants were evaluated:

1. Original configuration: 1×1 convolution, no bias, orthogonal initialization, and batch normalization.
2. Modified configuration: adding bias terms while removing batch normalization.

This comparison isolates the impact of output layer design on loss function performance, with the network producing single-channel probability maps representing fluorophore localization likelihood from single-channel (grayscale) microscopy images.

.

3.4.3 Full Dataset Training

The two superior performing models (Focal Loss with original and modified architectures) were trained on the complete 10,000-sample dataset for 20 epochs. This full-scale training validated findings from micro-subset analysis and provided final performance benchmarks for practical implementation.

3.4.4 Focal Loss Hyperparameters

Focal Loss introduces two key hyperparameters: α (alpha), which balances the weight of positive and negative examples, and γ (gamma), which adjusts the rate at which easily classified examples are down-weighted, thereby focusing training on harder, misclassified examples. While these parameters are crucial for optimizing performance, a systematic investigation of hyperparameter sensitivity and adaptive tuning strategies, such as grid search, was beyond the scope of this study. This decision was primarily driven by the computational intensity and significant time required for training these deep learning models on the large 10,000-sample dataset over 20 epochs. The determination of optimal values for α and γ, which may depend on specific imaging conditions, sample characteristics, and network architecture, remains a crucial area for future work.

3.5 Evaluation Framework

Quantitative Metrics: All models were evaluated using Peak Signal-to-Noise Ratio (PSNR), Structural Similarity Index (SSIM), and Mean Squared Error (MSE) for comprehensive image quality assessment. Localization Analysis: The two best-performing models underwent detailed Intersection over Union (IoU) analysis at threshold 0.1, with results presented using violin plot visualization to demonstrate performance distributions and statistical significance.

Visual Assessment: Systematic qualitative evaluation compared input diffraction-limited images, ground truth localizations, and model outputs across all training stages, enabling identification of artifacts, localization precision, and reconstruction quality differences between approaches.

IV. Results

4.1 Qualitative Performance Assessment

4.1.1 Original Architecture Loss Function Comparison

Systematic evaluation (Table 1.) revealed fundamental differences in reconstruction capabilities across loss functions. **L1+L2 and MSE losses** failed to converge to meaningful solutions(Fig. 3.), producing noisy, distorted versions of original input images with complete failure to localize emitters. **BCE loss** showed partial promise with large bright spots appearing over emitter locations but suffered from substantial noise and ghosting artifacts that preserved diffraction-limited image characteristics (Fig. 4). **Focal Loss** successfully localized multiple emitters with reasonable confidence and significantly reduced noise levels compared to other approaches (Fig. 4).

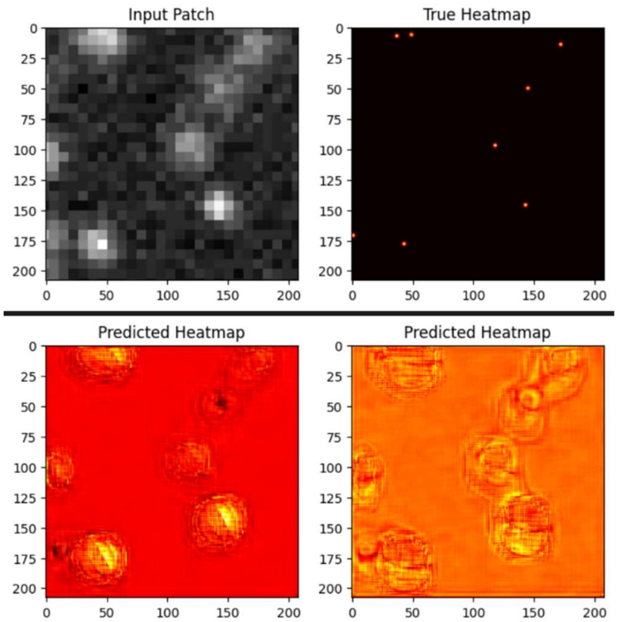

Fig. 3. Original Model training results, L1+L2 (left) and MSE (right)

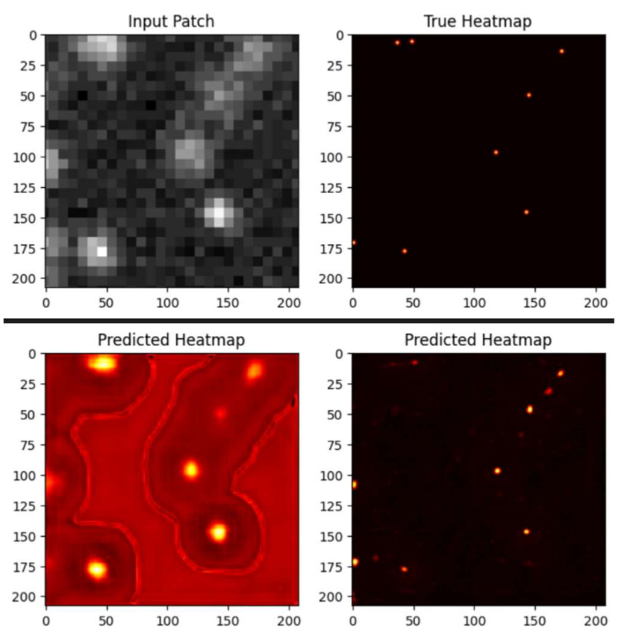

Fig. 4. Original Model Training results, BCE (left) and Focal Loss (right)

4.1.2 Modified Architecture Impact

Output layer modification (adding bias, removing batch normalization) maintained the same fundamental performance hierarchy with one notable exception: focal loss demonstrated small but noticeable improvements in reconstruction quality, particularly in emitter definition and noise reduction (Fig. 5).

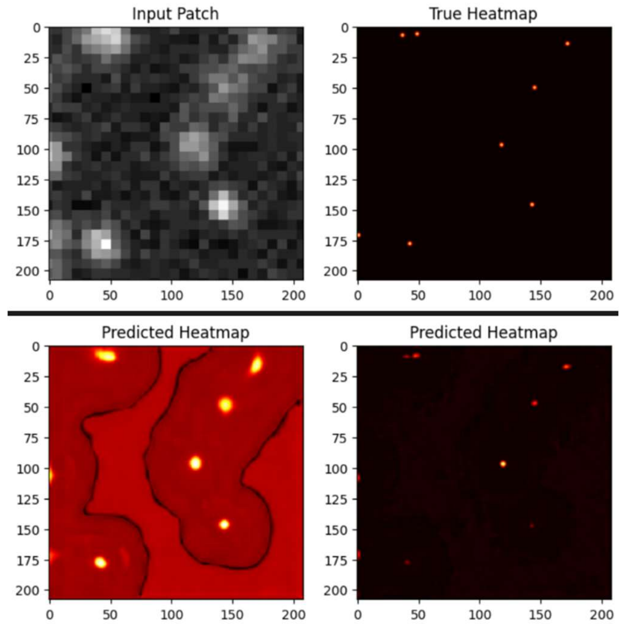

Fig. 5. Modified Model training results, BCE (left) and Focal Loss (right)

4.1.3 Full Dataset Training Results

Both focal loss variants trained on the complete dataset (10,000 samples) achieved exceptional reconstruction quality. The modified architecture version showed superior performance in emitter circularity and close-emitter distinction capabilities, demonstrating the combined benefits of optimal loss function selection and architectural refinement.

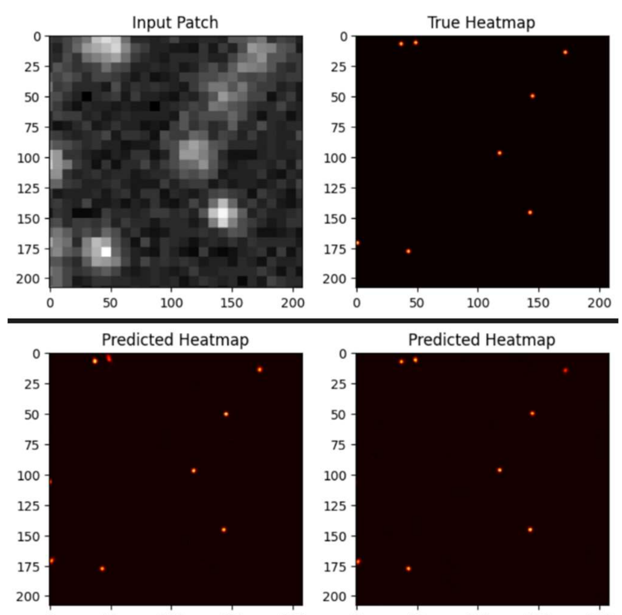

Fig. 6. Full Dataset Training results, Focal Loss before (left) and after (right) modification

4.2 Quantitative Performance Metrics

Comprehensive evaluation using PSNR, SSIM, and MSE metrics confirmed qualitative observations and established clear performance hierarchies.

Table 1. Models' Quantitative Evaluation of the Models

Loss-Function (original output layer	Evaluation metrics		
	PSNR	SSIM	MSE
Combination L1L2	20.1792	0.536008	0.010108
MSE	17.118616	0.076522	0.020488
BCE	-8.251812*	-0.000403*	227.16954*
Focal-Loss	31.905282	**0.833705**	0.000672
Output layer with Bias and no BatchNorm			
Combination L1L2	15.108538	0.005722	0.034216
MSE	19.711771	0.232899	0.011645
BCE	-9.294546*	-0.000333*	1871.049194*
Focal-Loss	31.193459	0.53439	0.000794
Model Trained on Full Dataset			
Focal-Loss	33.826982	0.260181	0.00043
Focal-Loss (modified)	33.414322	0.246735	0.000475

**BCE metrics show anomalous values due to binary-continuous output conflicts.*

4.3 Localization Precision Analysis

Intersection over Union (IoU) evaluation at threshold 0.1 (Fig. 7.) provided quantitative assessment of spatial localization accuracy for the two superior models:

- **Focal Loss (Modified Architecture)**: IoU = 0.500
- **Focal Loss (Original Architecture)**: IoU = 0.474

The modified architecture achieved 5.5% improvement in localization precision, validating the qualitative observations of enhanced emitter definition and spatial accuracy.

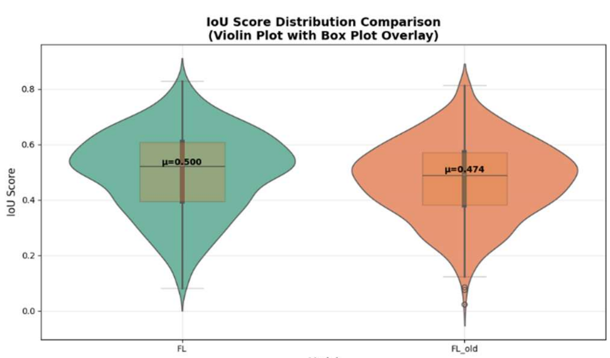

Fig. 7. IoU Violin plots, before (left) and after (right) modification

V. Discussion

5.1 Comprehensive Interpretation of Loss Function Performance

The results of the model evaluation (Table 1.) establish focal loss as categorically superior across all evaluation dimensions: only focal loss achieved meaningful localization solutions with PSNR >31 dB versus <20 dB for functional alternatives, SSIM 0.83 versus maximum 0.54 for other approaches, reconstruction fidelity MSE 0.0007 versus >0.01 for traditional losses, and 5.5% IoU improvement with architectural optimization. These quantitative metrics confirm that focal loss produces sharp, accurate localizations with minimal artifacts while traditional regression losses fail entirely for SMLM reconstruction tasks.

The failure of L1+L2 and MSE losses to reach competitive reconstruction quality highlights their incompatibility with the sparse, binary nature of SMLM. Both losses treat the task as continuous regression, leading to diffuse "contrast maps" rather than precise localizations. Although PSNR values of ~17–20 dB indicate some ability to capture broad intensity distributions, the reconstructions lack spatial precision necessary for accurate fluorophore localization.

Binary Cross-Entropy (BCE) fails dramatically, with negative PSNR values and unbounded MSE, producing unusable outputs. This represents fundamental breakdown: BCE's uniform weighting cannot cope with extreme class imbalance where <1% of pixels contain true emitters. Instead of precise reconstructions, BCE degenerates to pathological predictions.

5.1.1 Focal Loss: Addressing Fundamental Problem Characteristics

Focal loss achieves superior results with PSNR up to 33.8 dB, SSIM up to 0.83, and MSE as low as 4.3×10^{-4}. These represent 60–100% PSNR improvements over regression baselines. Its ability to down-weight abundant background pixels and emphasize difficult positives produces reconstructions with sharp, accurate localizations. IoU evaluation confirms this, with the best focal loss model achieving ~0.50 IoU@0.1 versus ~0.47 for baseline focal loss.

5.2 Implications and Significance for the Field

These findings necessitate reconsideration of optimization strategies in computational microscopy. Performance differences (PSNR 17–20 dB with regression losses vs. >31 dB with focal loss) demonstrate that architectural innovations alone are insufficient: loss function design is equally critical.

The superior performance of focal loss has immediate implications for practitioners. Reconstructions are sharp, consistent, and free from pathological failures of BCE and regression-based losses. The robustness and convergence reliability make focal loss a practical standard for super-resolution SMLM applications.

5.3 Critical Analysis of Architectural Modifications

Removing batch normalization and adding bias to the output layer produced mixed effects: regression-based losses improved slightly (MSE PSNR rose from 17.1 → 19.7 dB), while focal loss performance decreased somewhat (PSNR 31.9 → 31.2, SSIM 0.83 → 0.53). This suggests architectural tweaks have secondary importance compared to loss design.

The mixed effects likely arise because regression-based losses benefit from the added flexibility of bias terms when batch normalization is removed, allowing better pixel-wise fitting. In contrast, focal loss, which emphasizes hard-to-predict emitters, may rely on the stabilizing effect of batch normalization, so its removal reduces convergence stability and overall reconstruction quality.

5.4 Limitations and Critical Challenges

5.4.1 Synthetic Data Dependence

[5] The investigation relies exclusively on ThunderSTORM-generated synthetic data, which may not capture real experimental conditions complexity. Systematic evaluation on experimental datasets with varying noise characteristics, photobleaching dynamics, and sample heterogeneity remains necessary to validate generalizability.

Such experimental factors can degrade localization accuracy and convergence, meaning optimal loss functions and architectures identified on synthetic data may not generalize directly to real-world imaging.

5.4.2 Limited Architectural Exploration

While this study establishes focal loss superiority within the Deep-STORM U-Net framework, interaction between loss function selection and alternative architectures remains unexplored. The optimal loss function may vary with architectural complexity and design philosophy.

5.4.3 Computational Considerations

The investigation does not address computational efficiency differences between loss functions during training and inference. Comprehensive deployment requires analysis of training times, memory requirements, and inference speeds across different hardware configurations.

5.5 Future Research Directions

5.5.1 Experimental Validation and Robustness

Future work should systematically evaluate focal loss performance across diverse experimental datasets, including different fluorophore types, imaging modalities (STORM, PALM, DNA-PAINT [4]), and sample preparations, particularly under challenging conditions such as high background noise, dense labelling, and three-dimensional imaging.

5.5.2 Loss Function Innovation

The success of focal loss suggests promising avenues for developing specialized optimization strategies adapted to SMLM characteristics. In related domains like microscopy super-resolution, **hybrid loss functions**—combining pixel-wise reconstruction, perceptual, and adversarial losses—have been shown to significantly enhance fine-detail reconstruction over standard losses alone [10]. Similarly, in medical imaging contexts such as diffusion MRI super-resolution, **uncertainty-aware models** leveraging heteroscedastic noise modelling and Bayesian inference have successfully quantified predictive uncertainty and improved performance [11].
These precedents support the premise that **integrating perceptual, adversarial, or uncertainty components into focal-based losses** may yield both high-fidelity reconstructions and meaningful confidence estimates in deep learning–based SMLM.

.

5.5.3 Multi-Modal and Temporal Extensions

Extension to multi-colour SMLM applications and temporal dynamics (live-cell imaging) represents natural progressions. The optimization insights may prove crucial for handling increased complexity and data volumes associated with advanced imaging modalities.

5.5.4 Integration with Emerging Technologies

The focal loss optimization strategy should be evaluated within emerging deep learning paradigms, including diffusion models for super-resolution microscopy, physics-informed neural networks, and self-supervised learning approaches.

5.6 Broader Impact on Computational Microscopy

This investigation establishes a new standard for evaluating deep learning approaches in super-resolution microscopy, demonstrating that optimization strategy selection requires equal consideration with architectural innovation. The insights extend to any computer vision task involving sparse, binary detection in noisy environments, including astronomy (point source detection), medical imaging (lesion detection), and industrial inspection (defect identification). The substantial performance improvements achieved through principled loss function selection offer a pathway toward more reliable, accurate, and widely adopted computational super-resolution methods that can accelerate biological research across diverse applications.

VI. Conclusion and Future Work Considerations

6.1 Comprehensive Summary of Project Work

This research systematically investigated the critical role of **loss function selection** in deep learning-based single-molecule localization microscopy (SMLM), addressing a fundamental gap in computational super-resolution microscopy literature. By developing simulation tools, modernizing deprecated codebases, and designing controlled experiments, this study established **loss function optimization as equally important as architectural innovation** in determining reconstruction quality.
The investigation followed a two-stage experimental protocol: initial overfitting analysis on micro-subsets to isolate loss function behaviour, followed by full-scale training and evaluation using **10,000 ThunderSTORM-generated samples** [5]. Four loss functions were compared—combined L1+L2, Mean Squared Error (MSE), Binary Cross-Entropy (BCE), and Focal Loss—within both the original and modified Deep-STORM U-Net architectures [1].

6.2 Key Results and Critical Implications

The experiments revealed a clear **performance hierarchy** among loss functions:

- **L1+L2 and MSE**: produced PSNR values of ~15–20 dB and SSIM ≤0.53, failing to converge to precise localization solutions. Their outputs resembled contrast-enhanced diffraction-limited images rather than sparse emitter maps.
- **BCE**: failed catastrophically, with negative PSNR values (–8 to –9 dB) and extremely high MSE, indicating fundamental incompatibility with the extreme class imbalance of SMLM.
- **Focal Loss**: achieved categorical superiority, with PSNR up to **33.8 dB**, SSIM up to **0.83**, and MSE as low as $\mathbf{4.3\times10^{-4}}$. IoU@0.1 reached **0.50**, a **5% relative improvement** over the previous focal loss baseline (0.47).

This establishes focal loss as the only tested strategy capable of producing accurate, high-resolution reconstructions.

The results highlight that SMLM reconstruction cannot be optimized with general-purpose losses. Regression losses (L1+L2, MSE) mischaracterize the problem as continuous intensity prediction, while BCE collapses under severe class imbalance. **Focal Loss directly addresses these challenges** by reducing the dominance of abundant background pixels and emphasizing rare, informative emitter locations. This problem-specific alignment explains its superior precision and reliability.

These findings demand a paradigm shift in computational microscopy: **optimization design is as critical as network architecture**. While architectural changes (e.g., removing batch normalization, adding bias) provided modest effects, they were secondary to loss function choice. The insights generalize to other sparse detection domains—astronomy (star detection), medical imaging (lesions), and industrial inspection (defects)—where focal loss or related formulations may provide similar benefits.

6.3 Critical Discussion of Limitations

6.3.1 Synthetic Data Constraints

The exclusive reliance on ThunderSTORM-generated synthetic data represents the most significant limitation of this investigation [5]. While synthetic data provides controlled conditions essential for isolating loss function effects, real experimental data introduces complexities not captured in simulation: heterogeneous noise characteristics, photobleaching dynamics, sample drift, and biological variability. The generalizability of focal loss superiority across diverse experimental conditions remains to be definitively established.

6.3.2 Architectural Scope Limitations

[1] The investigation focused exclusively on the Deep-STORM U-Net architecture, leaving unexplored the interaction between loss function selection and alternative network designs. [2] Advanced architectures such as DECODE's dual U-Net approach, transformer-based methods, or physics-informed neural networks may exhibit different sensitivities to loss function choice, potentially altering the performance hierarchy observed in this study.

6.3.3 Evaluation Metric Dependencies

While comprehensive, the evaluation framework relies primarily on pixel-based metrics (PSNR, SSIM, MSE) and spatial overlap measures (IoU). These metrics may not fully capture biologically relevant performance characteristics such as molecular counting accuracy, spatial resolution limits, or robustness to imaging artifacts—factors critical for practical SMLM applications.

6.4 Insightful Recommendations for Future Work

6.4.1 Experimental Rigor and Methodological Improvements

Priority 1: Enhancement of experimental rigor through systematic training optimization strategies:

To enhance experimental rigor, several methodological improvements should be considered. **Adaptive learning rate scheduling** techniques, such as cosine annealing, step decay, or plateau reduction, could be employed to ensure optimal convergence across all loss functions. In addition, a **systematic evaluation of emitter densities** across a broader range (0.1–5.0 μm^{-2}) would help establish how loss function performance varies under different sparsity conditions. Robustness could be further strengthened through the use of **k-fold cross-validation protocols**, ensuring statistical significance of observed differences. Finally, **systematic hyperparameter optimization**—for example, using grid search or Bayesian optimization to tune focal loss parameters (α, γ)—would provide a more comprehensive understanding of how imaging conditions influence model performance.

Priority 2: Experimental data integration and validation:

Future work should also prioritize **direct training and testing on experimental datasets**, including structures such as microtubules, actin, and DNA origami. To bridge the gap between synthetic and experimental conditions, **domain adaptation techniques**, such as adversarial training, could be employed to improve generalization. In addition, **comparative evaluations against established localization software**—including QuickPALM, rapidSTORM, and SMAP—on standardized benchmarks would provide a rigorous baseline for assessing the practical impact of deep learning approaches.

Priority 3: Advanced evaluation frameworks:

Future evaluation frameworks should move beyond pixel-level measures to include **localization precision metrics** such as Nearest Neighbour Analysis (NeNA) and Fourier Ring Correlation (FRC) for more accurate resolution assessment. Equally important is the measurement of **molecular counting accuracy**, quantifying detection efficiency and false positive rates across different emitter density regimes. For live-cell applications, **temporal consistency** must be evaluated to ensure frame-to-frame stability of localizations. Finally, **biological relevance metrics** should be incorporated to verify that reconstructions preserve cellular structure integrity while supporting quantitative biological measurements.

6.4.2 Advanced Loss Function Innovation

The success of focal loss opens opportunities for developing even more specialized optimization strategies:

- **Hybrid focal-perceptual losses** combining pixel-level accuracy with feature-level similarity
- **Uncertainty-aware losses** providing confidence estimates for individual localizations
- **Physics-informed losses** incorporating known PSF characteristics and imaging physics
- **Multi-scale losses** optimizing reconstruction quality across different spatial frequencies

6.4.3 Architectural Integration and Optimization

Future work should systematically explore loss function performance across diverse network architectures:

- Integration with DECODE's temporal context framework [2]
- Application to transformer-based super-resolution methods
- Evaluation within physics-informed neural network approaches
- Extension to generative models (diffusion models, GANs) for SMLM reconstruction

6.4.4 Multi-Dimensional Extensions

Temporal dynamics: Extension to live-cell SMLM applications requiring optimization across both spatial and temporal dimensions, potentially through specialized loss functions incorporating motion models and temporal consistency constraints.

Multi-colour imaging: Development of loss functions optimized for simultaneous multi-target localization, addressing challenges of spectral crosstalk and differential labelling efficiency.

Three-dimensional reconstruction: Investigation of focal loss performance in 3D SMLM applications, including specialized loss formulations for depth-dependent PSF variations and axial localization accuracy.

6.5 Unresolved Questions and In-Depth Discussion

6.5.1 Fundamental Theoretical Questions

Why does focal loss succeed where others fail? While this investigation empirically demonstrates focal loss superiority, the theoretical foundations remain incompletely understood. Future work should develop mathematical frameworks explaining why focal loss alignment with SMLM characteristics leads to convergence success, potentially through information-theoretic analysis or optimization landscape characterization.

What are the fundamental limits of loss function optimization? The dramatic improvements achieved through focal loss raise questions about theoretical performance bounds. Recent work has demonstrated a model-free approach to estimate the Cramér–Rao bound for deep-learning microscopes and shown that convolutional neural networks can approach this ultimate precision limit for localization tasks [9]. Applying such theoretical frameworks to SMLM could help define achievable accuracy ceilings and inform future loss-function design and optimization.

6.5.2 Practical Implementation Challenges

Hyperparameter sensitivity: The focal loss formulation introduces additional hyperparameters (α, γ) whose optimal values may depend on imaging conditions, sample characteristics, and network architecture. Systematic investigation of hyperparameter sensitivity and adaptive tuning strategies represents crucial future work.

Computational efficiency trade-offs: While focal loss achieves superior reconstruction quality, the computational implications during training and inference require comprehensive analysis. The relationship between reconstruction accuracy and computational cost across different hardware configurations remains unexplored.

6.5.3 Broader Scientific Impact Questions

Transferability to adjacent fields: The success of focal loss in SMLM raises questions about applicability to related sparse detection problems in astronomy, medical imaging, and industrial inspection. Systematic investigation of transfer learning strategies could establish the broader impact of these optimization insights.

Integration with emerging technologies: How will focal loss optimization perform within next-generation approaches such as AI-designed microscopes, real-time adaptive imaging systems, or quantum-enhanced microscopy? These questions become increasingly relevant as the field advances toward fully automated, AI-driven imaging pipelines.

6.6 Final Synthesis

This investigation establishes loss function selection as a critical determinant of success in deep learning-based SMLM, with implications extending throughout computational microscopy and sparse detection applications. The systematic demonstration of focal loss superiority provides immediate practical guidance while revealing fundamental principles about optimization design for imbalanced detection tasks. The work opens multiple avenues for future research that could further revolutionize computational approaches to biological imaging, ultimately accelerating scientific discovery through more accurate, reliable, and widely accessible super-resolution microscopy tools.

The comprehensive nature of these findings, coupled with honest acknowledgment of limitations and clear directions for future work, positions this research as a foundational contribution that will guide optimization strategy development in computational microscopy for years to come.

VII. AI Ethical Considertions

Potential Biases and Data Representation

[5] The exclusive reliance on ThunderSTORM-generated synthetic training data introduces simulation bias, where models may perform well on idealized data but struggle with real experimental conditions. This could inadvertently favour certain experimental setups while disadvantaging underrepresented scenarios. Additionally, the superior performance of focal loss could create methodology bias, where researchers preferentially adopt this approach without considering limitations or alternative methods more suitable for specific applications.

Importantly, real-world SMLM datasets often involve animal or human tissue or cellular material. The use of such biological samples is subject to formal ethical approval and regulatory oversight. In this context, simulation tools become particularly valuable, as they can reduce dependence on animal or human material, aligning with the principles of the NC3Rs (Replacement, Reduction, and Refinement) promoted by the UK's NC3Rs initiative [NC3Rs, 2019]. Improvements in simulation quality therefore not only enhance methodological rigor but also support more ethically sustainable research practices.

Mitigation strategies include diversifying synthetic data parameters, systematic validation across real-world datasets, and transparent communication of focal loss limitations to maintain methodological diversity in computational microscopy.

Privacy Concerns and Data Security

While microscopy data may appear less sensitive than clinical data, **biological privacy concerns** remain relevant for proprietary research involving novel experimental techniques or competitive areas. Deep learning models might inadvertently encode information about experimental protocols that could be reverse engineered. **Model security** issues include potential malicious modification of training procedures, leading to systematically biased results with significant downstream scientific impacts.

Protection measures require secure data handling protocols, cryptographic verification of model integrity, and transparent documentation of training procedures.

Fairness and Equitable Access

The computational requirements for deep learning SMLM create **digital divide concerns**, where well-resourced institutions gain disproportionate access to advanced reconstruction capabilities. Additionally, **technical expertise barriers** limit access for researchers without extensive machine learning backgrounds.

Democratization efforts should prioritize computationally efficient models, cloud-based resources, pre-trained model availability, and comprehensive educational materials to ensure broad accessibility across the scientific community.

Transparency and Scientific Integrity

The "black box" nature of deep learning poses **interpretability challenges** in scientific applications where understanding model reasoning is crucial for validating results. The dramatic performance improvements achieved through focal loss demand robust evidence that gains reflect genuine advancement rather than overfitting to specific datasets.

Requirements include comprehensive documentation, public code/data availability, independent replication across research groups, and systematic validation protocols ensuring model predictions align with biological principles and experimental expectation.

References

E. Nehme, L. E. Weiss, T. Michaeli, and Y. Shechtman, “Deep-STORM: super-resolution single-molecule microscopy by deep learning,” *Optica*, vol. 5, no. 4, p. 458, Apr. 2018, doi: 10.1364/optica.5.000458.

Z. Chen *et al.*, “DECODE: a Deep-learning framework for Condensing enhancers and refining boundaries with large-scale functional assays,” *Bioinformatics*, vol. 37, no. Supplement_1, pp. i280–i288, Apr. 2021, doi: 10.1093/bioinformatics/btab283.

A. Speiser *et al.*, “Teaching deep neural networks to localize single molecules for super-resolution microscopy,” *arXiv.org*, Jun. 27, 2019. https://arxiv.org/abs/1907.00770

J. Schnitzbauer, M. T. Strauss, T. Schlichthaerle, F. Schueder, and R. Jungmann, “Super-resolution microscopy with DNA-PAINT,” *Nature Protocols*, vol. 12, no. 6, pp. 1198–1228, May 2017, doi: 10.1038/nprot.2017.024.

M. Ovesný, P. Křížek, J. Borkovec, Z. Švindrych, and G. M. Hagen, “ThunderSTORM: a comprehensive ImageJ plug-in for PALM and STORM data analysis and super-resolution imaging,” *Bioinformatics*, vol. 30, no. 16, pp. 2389–2390, Apr. 2014, doi: 10.1093/bioinformatics/btu202.

D. Palounek *et al.*, “Surpassing the diffraction limit in Label-Free optical microscopy,” *ACS Photonics*, Aug. 2024, doi: 10.1021/acsphotonics.4c00745.

“The diffraction barrier in optical microscopy,” *Nikon’s MicroscopyU*. https://www.microscopyu.com/techniques/super-resolution/the-diffraction-barrier-in-optical-microscopy

B. Huang, H. Babcock, and X. Zhuang, “Breaking the diffraction barrier: Super-Resolution imaging of cells,” *Cell*, vol. 143, no. 7, pp. 1047–1058, Dec. 2010, doi: 10.1016/j.cell.2010.12.002.

“Model-free estimation of the Cramér-Rao bound for deep-learning microscopy in complex media.” https://arxiv.org/html/2505.22330v1

“Reconstructing Interpretable Features in Computational Super-Resolution microscopy via Regularized Latent Search.” https://arxiv.org/html/2405.19112v1

R. Tanno *et al.*, “Bayesian Image Quality Transfer with CNNs: Exploring Uncertainty in dMRI Super-Resolution,” *arXiv.org*, May 01, 2017. https://arxiv.org/abs/1705.00664