# Super-resolution benchmarking for 3D image-to-image fusion problem

*Daniel Franco-Barranco[1,2,3,*], Aitor González-Marfil [3,4], Albert Cardona[1,2], Arrate Muñoz-Barrutia[5,6], Ignacio Arganda-Carreras[3,4,7,8]*

[1] MRC Laboratory of Molecular Biology, University of Cambridge, Cambridge, UK
[2] Department of Physiology, Development and Neuroscience, University of Cambridge, Cambridge, UK
[3] Donostia International Physics Center (DIPC), San Sebastian, Spain
[4] Dept. of Computer Science and Artificial Intelligence, University of the Basque Country (UPV/EHU), San Sebastian, Spain
[5] Dept. de Bioingenieria, Universidad Carlos III de Madrid, Madrid, Spain
[6] Area de Bioingenieria, Instituto de Investigación Sanitaria Gregorio Marañón, Madrid, Spain
[7] IKERBASQUE, Basque Foundation for Science, Bilbao, Spain
[8] Biofisika Institute (CSIC-UPV/EHU), Leioa, Spain

2025 IEEE 22nd International Symposium on Biomedical Imaging (ISBI) | 979-8-3315-2052-6/25/$31.00 ©2025 IEEE | DOI: 10.1109/ISBI60581.2025.10981218

## ABSTRACT

Fluorescence microscopy faces challenges in resolution, phototoxicity, and anisotropic artifacts. The *Fuse My Cells* challenge, organized by France-BioImaging, aims to develop deep learning models that predict fused 3D volumes from single-view acquisitions, reducing phototoxic exposure while enhancing resolution. In this work, we benchmark state-of-the-art super-resolution models, including DFCAN, RCAN-3D, UNETR, and a 3D-adapted RCAN-it, evaluating their performance on the *Fuse My Cells* challenge dataset, which encompasses 802 3D light-sheet microscopy images. A novel training strategy prioritizing high-discrepancy regions optimizes efficiency and improves reconstruction accuracy. Our findings suggest that super-resolution models can not fully reconstruct the information on those image areas with minimum signal information. Code and documentation can be found at `https://github.com/danifranco/BiaPy`.

***Index Terms*—** Bioimage analysis, deep learning, image restoration.

## 1. INTRODUCTION

Fluorescence microscopy enables high-resolution 3D imaging of biological systems but faces challenges such as resolution limitations, phototoxicity, and anisotropic artifacts. Multi-view fusion techniques mitigate these issues by integrating complementary perspectives, yet they require prolonged photon exposure, increasing photobleaching and reducing cell viability. The *Fuse My Cells* challenge, organized by France-BioImaging, addresses this limitation by introducing a large and heterogeneous 3D dataset of light-sheet microscopy images. This dataset aims to advance the field by facilitating the development of novel algorithms capable of predicting fused 3D volumes from a single view, thereby enhancing resolution while minimizing phototoxic exposure.

Traditional 3D image restoration methods, including denoising and deconvolution, have been used to improve image quality. Techniques such as BM3D [1] reduce noise, while the Richardson-Lucy algorithm [2, 3] refines optical blur through iterative deconvolution. However, these methods struggle with anisotropic 3D artifacts and are computationally expensive on large datasets [4].

Deep learning has revolutionized image restoration, particularly through super-resolution (SR) methods that enhance low-quality images by recovering high-frequency details [5]. Supervised frameworks such as CARE [6] leverage paired training data, while self-supervised approaches like Noise2Void [7] eliminate the need for clean references. Hybrid models integrate physics-based priors with neural networks, exemplified by SPITFIR(e) [8] and GAN-based cross-modality SR techniques [9, 10]. Despite these advancements, challenges persist in balancing computational efficiency, generalization across imaging modalities, and structural fidelity.

SR networks are particularly suited for single-view 3D reconstruction, as they infer and restore missing structural details from sparse or anisotropic inputs [5]. Convolutional neural networks (CNNs) and transformers [11] have demonstrated superior performance in this domain. Advanced attention mechanisms, such as Fourier channel attention in deep Fourier channel attention network (DFCAN) [12] and residual channel attention in Residual Channel Attention Network (RCAN) [13], enhance feature representation, while UNEt TRansformers (UNETR) [14] extends transformer-based modeling to volumetric data.

Corresponding author: dfranco@mrc-lmb.cam.ac.uk

In this work, we benchmark four state-of-the-art SR architectures tailored for single-view 3D reconstruction: DFCAN, RCAN-3D, RCAN-it, and UNETR. Our study evaluates their ability to balance reconstruction accuracy, computational efficiency, and cross-modality generalization. By leveraging deep learning to recover missing volumetric information, we provide insights into advancing high-resolution, photon-efficient fluorescence microscopy.

## 2. METHOD

### 2.1. Challenge Dataset

In two-view light-sheet microscopy, a fused 3D image is typically generated by capturing the sample from two orthogonal perspectives. These views are spatially aligned within a common reference frame to produce a unified 3D reconstruction. The primary objective of this dataset is to facilitate the development of deep learning methods capable of predicting the fused 3D image using only a single-view input, eliminating the need for multiple acquisitions.

The dataset used in this study consists of previously unpublished 3D images acquired through a two-view light-sheet microscopy, encompassing diverse conditions, tissue types, imaging techniques, resolutions, and intensity ranges. It includes a total of $802$ 3D images: $401$ single-view raw inputs and $401$ corresponding fused outputs, which serve as ground truth.

To optimize computational efficiency within the time constraints of the challenge edition, a subset of images - specifically those from *Study 1*, as labeled by the organizers[1]- was preprocessed by removing the first and last slices containing only background information. This step reduced dataset size and training time without affecting relevant image content.

### 2.2. Super-resolution model benchmarking

In this study, we present a benchmarking analysis of SR models, evaluating both biomedical-specific approaches, including DFCAN [12], RCAN-3D [15], and UNETR [14, 16], as well as widely adopted SR methods such as RCAN [13, 17].

**DFCAN**. Leverages frequency content differences in the Fourier domain, enhancing the precision and efficiency of high-frequency information reconstruction compared to traditional spatial domain approaches [12]. Building on the channel attention mechanism of the RCAN [13], DFCAN introduces a novel Fourier domain attention mechanism, demonstrating superior performance in generating SR images of diverse biological structures. It also exhibits robustness under challenging conditions, such as low signal-to-noise ratios and high background fluorescence.

In this study, we extend DFCAN from a 2D to a 3D framework. Despite the predominantly isotropic nature of the images, we adopt an input size of $16 \times 256 \times 256$, guided by evidence on the benefits of anisotropic input sizes in 3D imaging [15].

**RCAN-it**. We extended the 2D architecture proposed in [17] to a 3D framework. In this experiment, we evaluated different input sizes, including $64 \times 64 \times 64$—maintaining a small input size consistent with the original work—and $16 \times 256 \times 256$, reflecting alternative input sizes explored with other architectures in this study. The optimal performance was achieved with the $16 \times 256 \times 256$ input size, and therefore, all reported results are based on this configuration.

**RCAN-3D**. The RCAN-3D architecture extends the original 2D RCAN [13] by incorporating 3D convolutional layers, enabling volumetric data processing and enhanced axial resolution. This architecture was selected based on its demonstrated potential for fluorescence microscopy applications in prior work [15]. We adopt the same architecture as outlined in [15], with a key modification: replacing the original ReLU activation with the SiLU activation, as suggested by RCAN-it [17], to improve model performance and stability. Additionally, we use an anisotropic input size of $16 \times 256 \times 256$, aligning with the optimal configuration proposed by the original authors. This approach aims to enhance the model's capacity to capture high-resolution spatial details while maintaining robustness across diverse fluorescence microscopy datasets.

**UNETR**. Building on the success of this architecture in the previous Light My Cells challenge, presented by France-BioImaging at ISBI 2024 [16], we applied the same architecture and pretraining strategy to the current task. For this experiment, we employed a cubic[2] input size of $160 \times 160 \times 160$, ensuring that the model captures sufficient information about the light degradation present in the samples across all dimensions.

### 2.3. Training with meaningful patches

Since the input image constitutes one of the two views that comprise the ground truth (fused image), certain regions of the fused image are predominantly composed of information from the input image. Consequently, these regions do not pose a significant challenge for the model, as most of the necessary information is already present. The difficulty of the task, therefore, lies in reconstructing the areas for which the input image provides no information. To address this, we developed a custom data loader that trains exclusively on regions exhibiting the highest discrepancies relative to the ground truth. This approach involves computing the absolute difference between the input image and the ground truth to identify the most informative regions. Additionally, patches corresponding to background areas are discarded to avoid unnecessary computation. These measures resulted in approximately a $50\%$ reduction in training time.

[1] https://seafile.lirmm.fr/published/fusemycells/

[2] In UNETR the input size must be square

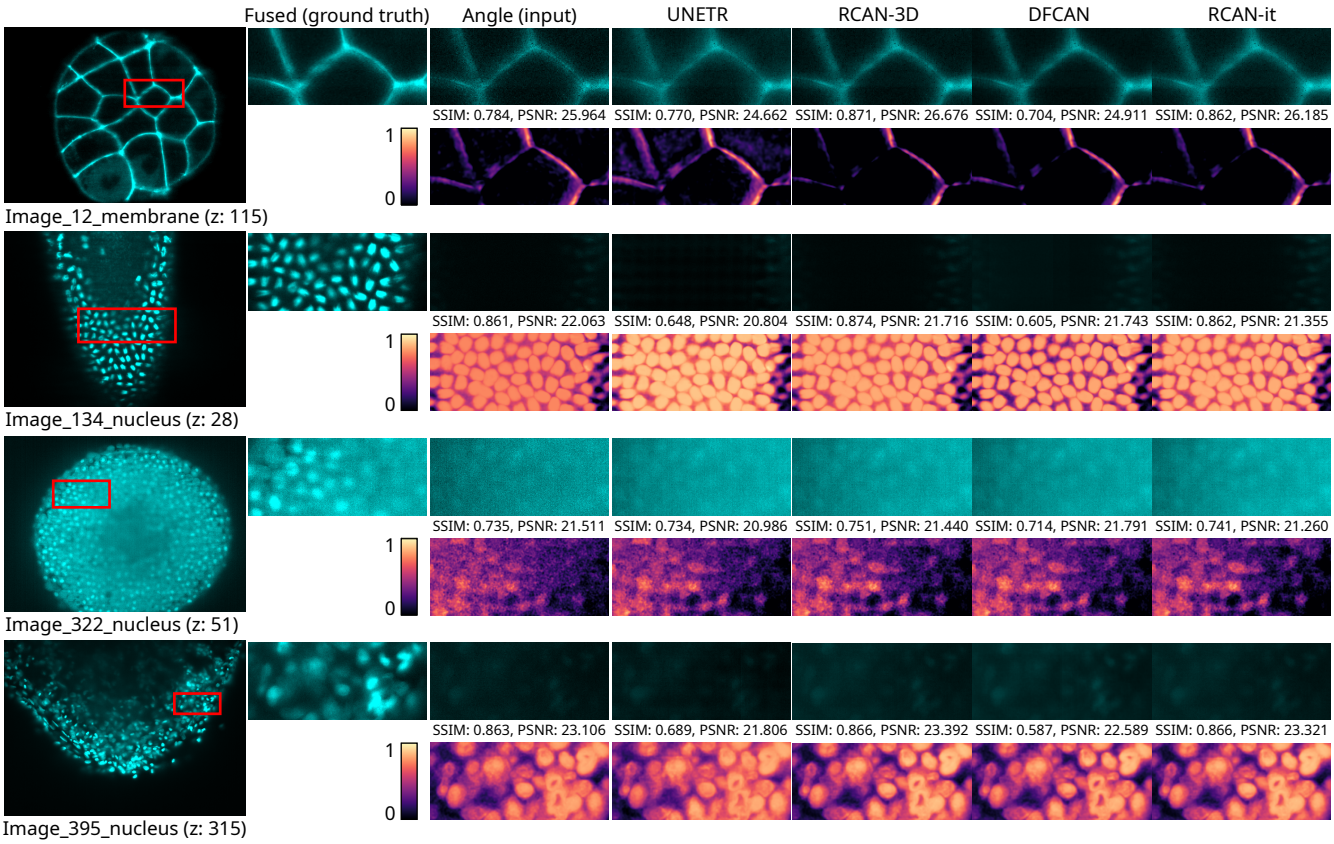


**Fig. 1**: Qualitative evaluation of the benchmarked models on the validation images. Each column represents a different model reconstruction. For each image, the first row displays a zoomed-in section of the image, corresponding to the red bounding box in the original input. The second row presents the inverted SSIM map ($1 - SSIM$), where brighter regions indicate higher discrepancies from the ground truth. SSIM and PSNR values are reported for each specific image. The red-marked regions were carefully selected to highlight areas with minimal information in the input image (angle), allowing an evaluation of the ability of each model to infer and reconstruct missing details in these low-content regions. This comparison provides insights into the effectiveness of SR networks in recovering structural fidelity from single-view inputs.

| Models | Metrics | | | |
|---|---|---|---|---|
| | PSNR ↑ | SSIM ↑ | MAE ↓ | Loss ↓ |
| Angle (input) | 23.112 | 0.7484 | **0.0456** | - |
| UNETR [14] | 22.169 | 0.6725 | 0.0562 | 0.1096 |
| RCAN-3D [15] | **23.310** | **0.7663** | 0.0460 | **0.0374** |
| DFCAN [12] | 21.751 | 0.6531 | 0.0666 | 0.0428 |
| RCAN-it [17] | 23.097 | 0.7646 | 0.0470 | 0.0380 |

**Table 1**: Aggregated results in the validation set of the different architectures benchmarked in the study. These values were computed by applying a percentile clipping at the 0.2th and 99.8th percentiles of the image intensities, following the guidelines provided by the challenge organizers. Best values marked in **bold**.

### 2.4. Implementation details

Our approach is fully integrated within the BiaPy library [18], using Pytorch version 2.4. We use basic data augmentation techniques that are commonly employed in SR workflows, such as flips and square rotations. We set aside 10% of the training samples for validation, which contains samples from all studies. A combination of mean absolute error (MAE) and Structural Similarity Index Measure (SSIM) was used as a loss function as in [12], specifically $0.8 * MAE + 0.2 * (1 - SSIM)$, with a learning rate of 0.0001, employing a cosine-decay scheduler with warm-up [19], and ADAMW optimizer. All models were trained for 100 epochs using eight A100 82GB GPUs with a patience of 10.

On top of that, to monitor the performance in the validation set, we also measure the Peak Signal-to-Noise Ratio (PSNR) and MAE, as shown in Table 1 and Figure 1.

### 2.5. Results on FuseMyCells challenge

As shown in Table 1, the results across all evaluated models are closely aligned, demonstrating minimal deviation from the baseline values obtained using the input images (angle column). While the SSIM indicates a potential improvement over the baseline, this enhancement appears to be marginal.

To provide a clearer understanding of the reconstruction capabilities of the networks with respect to the input images, Figure 1 was generated. This figure highlights regions with limited information in the input image (angle) where the networks must exert greater effort to reconstruct the signal. Despite the observed increase in SSIM values, as shown in Table 1, the qualitative assessment does not reveal an equally evident improvement to the human eye.

Overall, the high SSIM and PSNR values reported in Table 1 and Figure 1 can be attributed to the large volumetric nature of the images. Since the differences between the input and output images are primarily confined to regions lacking information in the original input/angle image, the quantitative metrics may not fully capture the perceptual quality of the reconstructed images.

## 3. CONCLUSION

In this study, we evaluated multiple super-resolution models within the FuseMyCells challenge. The results showed that all models performed similarly, with minimal deviation from the baseline established by the input images (Angle column). While the SSIM metric indicated a slight improvement, this enhancement was marginal and not consistently evident in the qualitative assessments.

Figure 1 highlighted the challenges in reconstructing regions with limited information, where the models needed to exert more effort. Although the high SSIM and PSNR values likely reflect the large volumetric size of the images, these metrics did not fully capture perceptual improvements, particularly in low-information areas.

Overall, while the models maintained baseline image quality, their potential for substantial perceptual enhancement was limited. Future research may explore several directions to improve performance:

- Incorporating generative models (e.g., diffusion models, generative adversarial networks [GANs]) to enhance the capacity for image synthesis and content generation.
- Revisiting the training strategy employed in this benchmark, which relied exclusively on patches where the difference between the input image and the ground truth exceeded a predefined threshold. This approach may not fully capture the underlying challenges of the problem. In certain cases, although the differences were substantial, the input signal remained strong, as observed in membrane structures and some nucleus images (see the first and third rows in Figure 1). Conversely, other regions contained little to no signal, such as the nuclei in the second and fourth rows of Figure 1. These observations suggest that the models failed to adequately differentiate between regions requiring reconstruction and those where the original information should be preserved with minimal modification. Consequently, this led to an overall averaging of the error across patches, preventing improvement beyond the observed performance plateau. So future work can explore advanced evaluation metrics and training strategy to better handle the reconstruction of low-information regions in volumetric imaging.
- In this study, we aimed to enhance all organelles simultaneously; however, developing more specialized models, one specifically targeting membranes and another focusing solely on the nucleus, may yield improved results.

**Acknowledgments.** This work is partially supported by grant GIU23/022 funded by the University of the Basque Country (UPV/EHU), and grants PID2021-126701OB-I00 and PID2023-152631OB-I00 funded by the Ministerio de Ciencia, Innovación y Universidades, AEI, MICIU/AEI/10.13039 /501100011033 and by "ERDF A way of making Europe".

**Compliance with Ethical Standards.** This work is a study for which no ethical approval was required.

## 4. REFERENCES

[1] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image Denoising by Sparse 3-D Transform-Domain Collaborative Filtering," *IEEE Transactions on Image Processing*, vol. 16, no. 8, pp. 2080–2095, 2007.

[2] W. H. Richardson, "Bayesian-based iterative method of image restoration∗," *J. Opt. Soc. Am.*, vol. 62, no. 1, pp. 55–59, Jan 1972.

[3] L. B. Lucy, "An iterative technique for the rectification of observed distributions," *Astronomical Journal, Vol. 79, p. 745 (1974)*, vol. 79, pp. 745, 1974.

[4] S. Preibisch, F. Amat, E. Stamataki, M. Sarov, R. H. Singer, E. Myers, and P. Tomancak, "Efficient bayesian-based multiview deconvolution," *Nature Methods*, vol. 11, no. 6, pp. 645–648, Jun 2014.

[5] Z. Wang, J. Chen, and S. C. H. Hoi, "Deep learning for image super-resolution: A survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 43, no. 10, pp. 3365–3387, 2021.

[6] M. Weigert, U. Schmidt, T. Boothe, A. Müller, A. Dibrov, A. Jain, et al., "Content-aware image restoration:

pushing the limits of fluorescence microscopy," *Nature Methods*, vol. 15, no. 12, pp. 1090–1097, Dec 2018.

[7] A. Krull, T-O. Buchholz, and F. Jug, "Noise2void - learning denoising from single noisy images," *arXiv preprint 1811.10980*, 2019.

[8] S. Prigent, H-N. Nguyen, L. Leconte, C. A. Valades-Cruz, B. Hajj, J. Salamero, and C. Kervrann, "SPIT-FIR(e): a supermaneuverable algorithm for fast denoising and deconvolution of 3D fluorescence microscopy images and videos," *Scientific Reports*, vol. 13, no. 1, pp. 1489, Jan 2023.

[9] H. Wang, Y. Rivenson, Y. Jin, Z. Wei, R. Gao, H. Günaydın, L. A. Bentolila, C. Kural, and A. Ozcan, "Deep learning enables cross-modality super-resolution in fluorescence microscopy," *Nature Methods*, vol. 16, no. 1, pp. 103–110, Jan 2019.

[10] P. Wijesinghe, S. Corsetti, D. J. X. Chow, S. Sakata, K. R. Dunning, and K. Dholakia, "Experimentally unsupervised deconvolution for light-sheet microscopy with propagation-invariant beams," *Light: Science & Applications*, vol. 11, no. 1, pp. 319, Nov 2022.

[11] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, et al., "An image is worth 16x16 words: Transformers for image recognition at scale," *arXiv preprint arXiv:2010.11929*, 2020.

[12] C. Qiao, D. Li, Y. Guo, C. Liu, T. Jiang, Q. Dai, and D. Li, "Evaluation and development of deep neural networks for image super-resolution in optical microscopy," *Nature Methods*, vol. 18, no. 2, pp. 194–202, 2021.

[13] Z. Yulun, L. Kunpeng, L. Kai, W. Lichen, Z. Bineng, and F. Yun, "Image super-resolution using very deep residual channel attention networks," *arXiv preprint*, p. 1807.02758, 2018.

[14] A. Hatamizadeh, Y. Tang, V. Nath, D. Yang, A. Myronenko, B. Landman, H. R Roth, and D. Xu, "UNETR: Transformers for 3D medical image segmentation," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV'22)*, Hawaii, USA, Jan 2022, pp. 574–584.

[15] J. Chen, H. Sasaki, H. Lai, Y. Su, J. Liu, Y. Wu, A. Zhovmer, C. A. Combs, I. Rey-Suarez, H.-Y. Chang, et al., "Three-dimensional residual channel attention networks denoise and sharpen fluorescence microscopy image volumes," *Nature Methods*, vol. 18, no. 6, pp. 678–687, 2021.

[16] D. Franco-Barranco, A. González-Marfil, and I. Arganda-Carreras, "Self-supervised vision transformers for image-to-image labeling: a BiaPy solution to the lightmycells challenge," in *2024 IEEE International Symposium on Biomedical Imaging (ISBI'24)*, Athens, Greece, May 2024, pp. 1–5.

[17] Z. Lin, P. Garg, A. Banerjee, S. A. Magid, D. Sun, Y. Zhang, L. Van Gool, D. Wei, and H. Pfister, "Revisiting RCAN: Improved training for image super-resolution," *arXiv preprint arXiv:2201.11279*, 2022.

[18] D. Franco-Barranco, J. A. Andres-San Roman, L. Hidalgo-Cenalmor, I.and Backova, A. Gonzalez-Marfil, C. Caporal, A. Chessel, P. Gomez-Galvez, L. M. Escudero, D. Wei, A. Muñoz-Barrutia, and A. Arganda-Carreras, "BiaPy: Accesible deep learning on bioimages," *bioRxiv*, p. 2024.02.03.576026, 2024.

[19] I. Loshchilov and F. Hutter, "Sgdr: Stochastic gradient descent with warm restarts," *arXiv preprint arXiv:1608.03983*, 2016.