From Simulation to Surgery: Enhancing Visual Realism in Minimally Invasive Mitral Valve Procedures

AI-enhanced image translation for clinically relevant endoscopic simulation in mitral valve repair.

Patrick Carnahan[1,2], Emma Zhang[2,3], John Moore[2], Terry M. Peters[2,4,5], Elvis C.S. Chen[2,4,5,6], Gianluigi Bisleri[1]

[1] Cardiac Surgery Department, St. Michael's Hospital, University of Toronto, Toronto, Canada;
[2] Imaging Laboratories, Robarts Research Institute, Western University, London, Canada;
[3] Schulich School of Medicine & Dentistry, Western University, London, Canada
[4] Department of Medical Biophysics, Western University, London, Canada;
[5] School of Biomedical Engineering, Western University, London, Canada;
[6] Electrical and Computer Engineering, Western University, London, Canada

St. Michael's
Inspired Care.
Inspiring Science.

Background

- Simulation-based training is increasingly important in competency-based medical education because it allows trainees to build technical skill, judgment, and procedural confidence in a controlled environment before entering the operating room [1].
- Minimally invasive mitral valve repair depends on an endoscopic view, so surgeons must interpret anatomy, depth, tissue appearance, and instrument interactions through the camera rather than direct visualization.
- Silicone mitral valve phantoms are practical for simulation-based training, but their visual appearance differs substantially from the intraoperative field.
- This mismatch may limit trainee immersion and reduce transfer of visual-anatomical skills from the simulator to the operating room.
- Prior work by Engelhardt *et al.* [2] showed that hyperrealistic image translation can improve mitral valve training phantoms, motivating further work on realism and temporal stability.

Figure 1: Silicone mitral valve phantom, endoscopic phantom view, and translated output demonstrating increased tissue coloration, blood-like appearance, and specular reflections.

Aim

- Develop and validate a surgically relevant image dataset and neural image translation framework to improve the visual realism of mitral valve training phantoms.
- Better replicate the live operative appearance of minimally invasive mitral valve surgery.
- Evaluate both perceptual realism and temporal stability, since useful simulation must support continuous endoscopic viewing rather than isolated still images.

Methods

- Collected endoscopic video data from minimally invasive mitral valve repair procedures and from silicone mitral valve phantoms positioned to match operative viewing angles.
- Curated 59,411 surgical frames from 570 surgical videos and 6,625 phantom frames from the simulator domain.
- Compared unpaired image-to-image translation models including CycleGAN, MUNIT, and DRIT [3].
- Tested patch-based consistency constraints to reduce frame-to-frame flickering during video playback.
- Evaluated perceptual similarity and temporal consistency in both static images and dynamic video sequences

Dataset and Workflow

- We built a surgical image dataset that captures clinically relevant mitral anatomy, endoscopic lighting conditions, tissue appearance, and instrument interactions seen during minimally invasive repair.
- Phantom images were collected under endoscopic viewpoints chosen to resemble the depth, angle, and field of view seen in the real surgical videos.
- Neural image translation was then applied to convert visually simple phantom images into outputs that more closely resembled the live operative field.

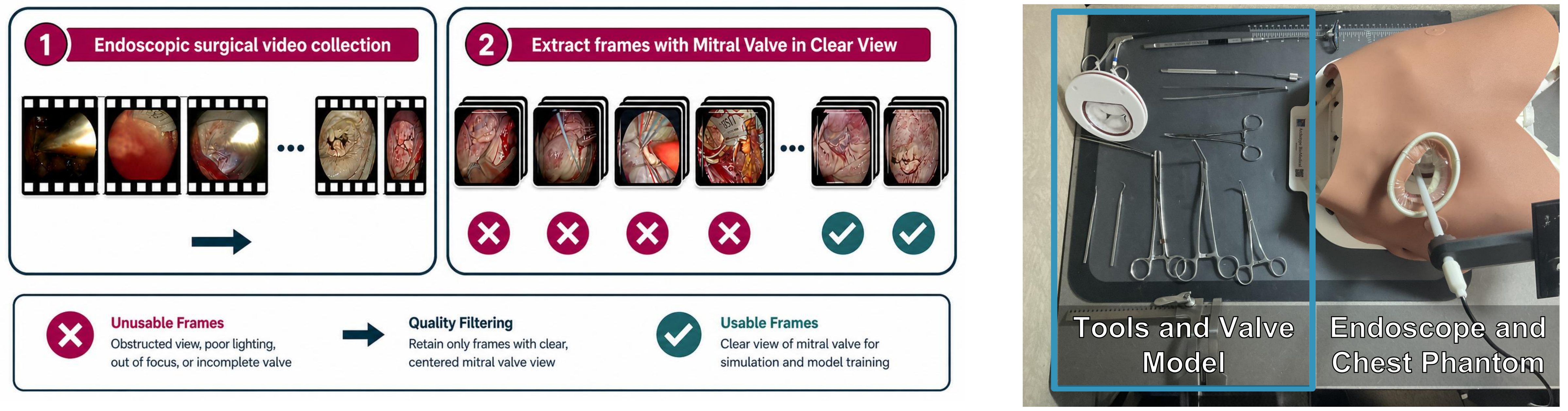

Figure 2: Workflow for surgical and phantom dataset collection, model training, and translated image generation.

Results

- Image translation improved phantom realism by adding more representative tissue texture, coloration, lighting effects, and instrument context to the simulator.
- Standard style translation produced acceptable realism in static images, but greater visual detail in video sequences was associated with increased frame-to-frame flickering, indicating a trade-off between realism and stability.
- Incorporating temporal consistency constraints reduced flicker and improved output stability while maintaining realism, supporting more continuous and usable endoscopic visualization for training.
- Increasing dataset size was associated with improved image fidelity and better visual continuity across adjacent frames in the translated sequences.

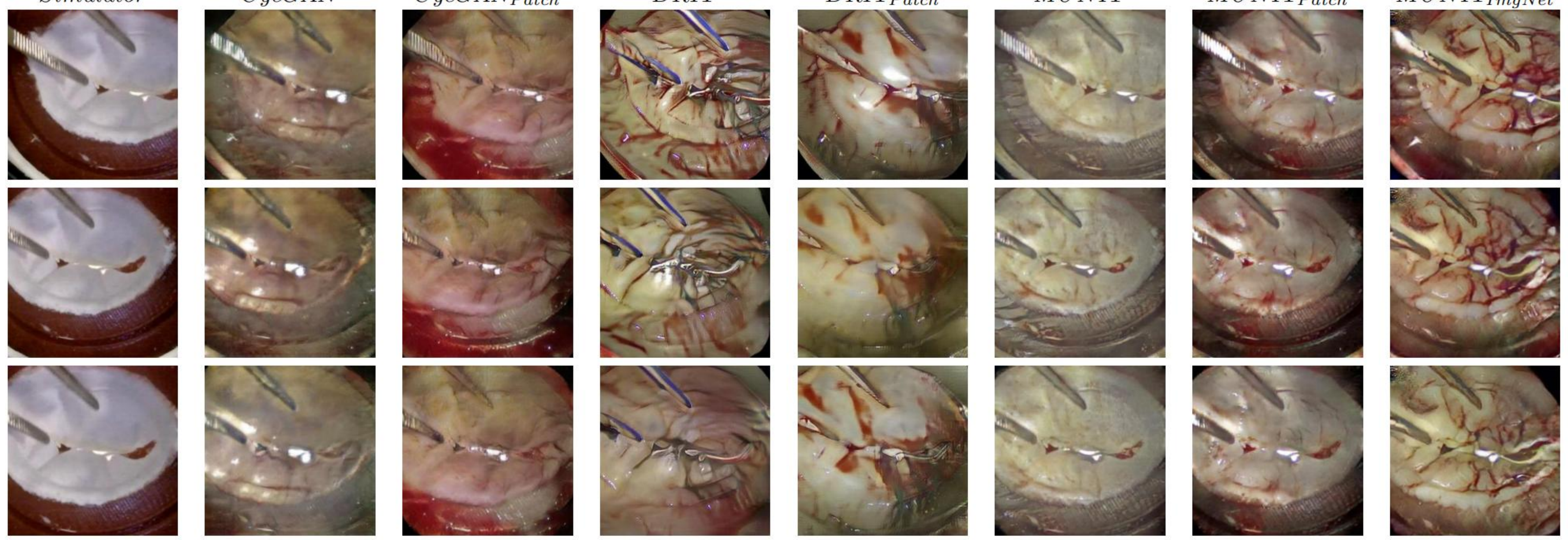

Figure 3: Representative model outputs showing improved realism and the trade-off between image detail and temporal stability. Each row is a different view, each column a different translation model. Ideal results would be realistic in appearance and consistent down columns.

Results (Cont'd)

- Image translation improved phantom realism, adding tissue coloration, specular highlights, and more operative-looking visual context.
- Best perceptual realism was achieved by **MUNITImgNet**, with perceptual similarity (FID) score **16.3 ± 3.2** versus **79.7 ± 17.9** for baseline simulator images.
- **CycleGANPatch** improved both realism and stability, reducing FID from **42.7 ± 9.1** to **30.2 ± 4.5** while also improving temporal consistency.
- Overall, lower FID scores generally came with greater frame-to-frame variation, highlighting a realism-versus-stability trade-off relevant to video-based simulation.

Perceptual Similarity and Patch Consistency Across Models

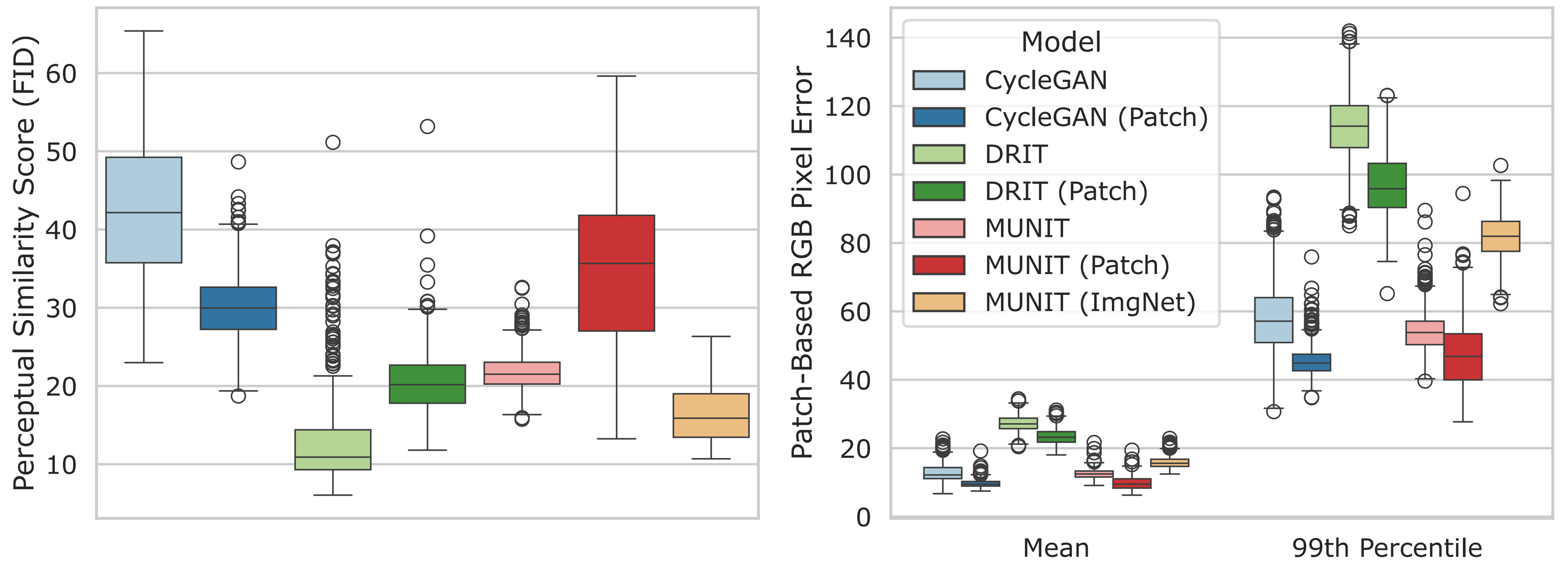

Figure 4: Quantitative comparison of perceptual similarity and temporal consistency across tested architectures. Lower is better across all metrics.

Conclusions

- AI-based image translation can narrow the visual gap between silicone mitral valve phantoms and the live operative field.
- For clinical simulation, photorealism alone is not enough; stable video output is also required for continuous endoscopic training.
- These findings support integration of image translation into next-generation minimally invasive and robotic mitral simulation platforms.
- Future work will focus on surgeon validation, trainee performance studies, and improved temporally consistent models for real-time deployment.

References

1. Frank JR, Snell LS, Cate OT, et al. Competency-based medical education: theory to practice. Med Teach. 2010;32(8):638-645.
2. Engelhardt S, DeSimone R, Full PM, Karck M, Wolf I. Improving surgical training phantoms by hyperrealism: deep unpaired image-to-image translation from real surgeries. In: Bildverarbeitung für die Medizin 2019. Wiesbaden, Germany: Springer Fachmedien Wiesbaden; 2019:282-282.
3. Zhang E, Rushlow A, Xia W, et al. Image style translation for improved imaging of mitral valve surgery training phantoms. In: Proc SPIE. 2025;13408:134080J. doi:10.1117/12.3046142.

Acknowledgements

NSERC CRSNG
CIHR IRSC
INNOVATION.CA
Western Biomedical Imaging Research Centre