
Enhancing Image-Guided Surgery with DL-Based 2D-3D X-Ray Registration



The integration of preoperative three-dimensional computed tomography (CT) and intraoperative two-dimensional X-ray imaging has significantly improved the precision of image-guided surgery. The ability to align these modalities enables more accurate lesion localisation, reduces surgical invasiveness and minimises the risk of damage to surrounding tissues. However, conventional registration methods face challenges in achieving the required speed and accuracy due to complex non-convex optimisation spaces and constrained search ranges. Recent advancements in deep learning have introduced robust solutions that enhance the efficiency of 2D-3D X-ray image registration. A novel approach using an enhanced Swin transformer framework demonstrates substantial improvements in accuracy, efficiency and robustness, offering promising applications in surgical navigation and radiotherapy.

Limitations of Traditional Image Registration Methods

Conventional methods for 2D-3D X-ray image registration rely on iterative optimisation techniques that refine pose estimations by minimising similarity differences between intraoperative X-rays and preoperative CT scans. Algorithms such as Powell's method and the covariance matrix adaptation evolution strategy (CMA-ES) are widely used for complex nonlinear registration problems. However, these approaches often suffer from slow computation times and high susceptibility to local minima, particularly when dealing with large capture ranges. The need for iterative refinement further increases computational complexity, making real-time applications impractical.

Precomputed feature extraction techniques, such as gradient-based volume analysis and spatial histograms, have been introduced to improve accuracy by reducing irrelevant information in the images. Despite these refinements, traditional methods remain limited in their ability to handle large-scale datasets and complex anatomical variations. As the capture range expands, the search space for optimisation grows, necessitating additional iterations to identify the optimal transformation, ultimately impacting both accuracy and speed. In clinical practice, these limitations create challenges in maintaining precise alignment during dynamic surgical procedures.

Deep Learning and the Swin Transformer Approach

Deep learning-based registration methods offer an alternative to traditional optimisation approaches by shifting computational complexity from the intraoperative phase to the preoperative phase. Instead of relying on iterative calculations, trained neural networks predict the optimal registration transformation in real-time based on learned patterns. Previous deep learning models, including convolutional neural networks (CNNs), have demonstrated promise in small-scale registration tasks but have struggled with global anatomical alignment due to limited semantic information capture.

A more advanced approach utilises the Swin transformer, which introduces a hierarchical structure with a shifted window mechanism. This design enables multi-scale feature extraction while maintaining computational efficiency. By combining attention mechanisms and a feature pyramid network, the Swin transformer improves both local and global feature representation, making it well-suited for medical image analysis.

The dual-channel X-ray Pose Estimation Swin Transformer (XPE-ST) applies this framework to the 2D-3D X-ray registration problem. Using digitally reconstructed radiographs (DRRs) generated from preoperative CT scans, the model is trained to estimate the six-degree-of-freedom pose alignment between X-ray and CT images. The incorporation of channel attention mechanisms and multi-stage feature fusion ensures that the extracted features maintain semantic consistency across different anatomical regions. As a result, XPE-ST significantly outperforms both traditional optimisation-based methods and other deep learning architectures, such as ResNet and DenseNet, in accuracy and efficiency.

Performance, Robustness and Clinical Applications

The XPE-ST model achieves a mean rotation error of 0.142° and a mean translation error of 0.362 mm, demonstrating a considerable improvement over existing methods. Unlike conventional optimisation-based approaches that require repeated DRR generation at multiple angles, XPE-ST reduces registration time to under 0.02 seconds, enabling real-time surgical applications.

Comparative experiments reveal that traditional optimisation methods exhibit substantial uncertainty and failure rates due to their sensitivity to anatomical variations. The likelihood of registration failure increases with capture range expansion, as optimisation algorithms struggle to converge in highly nonlinear search spaces. Deep learning-based methods offer greater resilience, but models such as ResNet and DenseNet remain sensitive to noise, affecting their performance in practical clinical settings.

XPE-ST maintains robust accuracy across different anatomical regions, including the head, chest and pelvis. Even under high levels of noise, which simulate real-world variations in X-ray quality, the model consistently achieves precise registration. The integration of a Swin transformer backbone allows for superior feature extraction, while attention mechanisms and multi-scale fusion improve the model's ability to adapt to varying conditions.

These improvements have significant implications for clinical applications. The ability to rapidly and accurately align preoperative CT scans with intraoperative X-rays enhances surgical navigation, enabling surgeons to make more informed decisions with greater confidence. In targeted radiotherapy, precise patient positioning ensures that radiation doses are delivered accurately to the treatment area while minimising exposure to healthy tissues. Additionally, in robot-assisted surgery, real-time registration supports precise robotic guidance, improving procedural outcomes.

Deep learning-based 2D-3D X-ray image registration presents a transformative advancement in medical imaging, addressing the speed and accuracy limitations of traditional optimisation methods. The XPE-ST model leverages the capabilities of the Swin transformer to provide a highly efficient and robust solution for image-guided interventions. Achieving both real-time processing speeds and high registration accuracy, this approach holds great potential for surgical navigation, radiotherapy and other clinical applications. Future developments will focus on incorporating real X-ray datasets, multi-view registration techniques and adaptations for non-rigid anatomical structures to further expand the model's applicability in medical imaging.

Source: [Bioengineering](#)

Image Credit: [iStock](#)

Published on : Sun, 2 Feb 2025