# Appendix 1: Search Strategy

## 1.1 Protocol Strategy

| | |
|---|---|
| Review Title: | **Deep Learning for Surgical Instrument Recognition and Segmentation in Robotic-Assisted Surgeries: A Systematic Review** |
| Objectives: | - Identifying all methods that are used for annotating tools in a robotic-assisted surgeries<br>- Classifying each model's features, use cases and performance.<br>- Evaluating accuracy, F1 score, and precision, of different deep learning Models in annotating different parts of minimally invasive surgeries. |
| Review question: | How can deep learning be used for surgical instrument recognition and segmentation in Robot-Assisted Surgery? |
| Inclusion criteria | **Minimally invasive surgery:**<br>• Adrenalectomy<br>• Brain surgery<br>• Colectomy<br>• Gallbladder surgery, also called cholecystectomy<br>• Heart surgery<br>• Kidney removal, also called radical nephrectomy<br>• Partial nephrectomy<br>• Kidney transplant<br>• Spine surgery<br>• Splenectomy<br>• MIS<br>• Endoscopy<br>• Laparoscopy<br>**Deep learning:**<br>• Convolutional Neural Networks (CNNs)<br>• Long Short Term Memory Networks (LSTMs)<br>• Recurrent Neural Networks (RNNs)<br>• Generative Adversarial Networks (GANs)<br>• Radial Basis Function Networks (RBFNs) |

| | |
|---|---|
| | • Multilayer Perceptrons (MLPs)<br>• Autoencoders<br>• Transfomers<br>**<u>Primary research</u>**<br>• Clinical trials<br>• Cohort<br>• RCT<br>• Report<br>• Case control study |
| Exclusion criteria | • Open Surgeries (Not minimally invasive)<br>• Laparoscopic Surgeries (not Robot Assisted)<br>• AI or ML but not deep learning Secondary research<br>• Systematic review<br>• Scoping review<br>• Meta analysis<br>• Literature review<br>• Review<br>• Response<br>• Letter<br>• Meta synthesis |
| Language | English |
| Databases | • PubMed<br>• IEEE Xplore<br>• Scopus<br>• EMBASE<br>• MEDLINE<br>• Web of Science |

### 1.2 Search Terms

| Functional Keyword | Text word | Controlled Vocabulary |
|---|---|---|
| Deep | Deep | Neural Networks, Computer* |

| | | |
|---|---|---|
| Learning | Learning | Deep Learning* |
| | | Image Processing, Computer-Assisted |
| | | Automation |
| Minimally Invasive Surgery | Minimally Invasive Surgical | Minimally Invasive Surgical Procedures |
| | | Laparoscopy |
| | | Surgical Instruments |
| | | Nephrectomy |
| | | Surgical Procedures, Operative |
| | | Surgery, Computer-Assisted |
| Annotation | Annotation | Data Annotation |
| | | Data Curation |
| | | Segementation |
| Tools | Tools | Surgical Instruments |

### 1.3 Concepts

**Review question:** *How can deep learning be used for surgical instrument recognition and segmentation in Robot-Assisted Surgery?*

Concept 1: Minimally Invasive Surgery

Keywords: laparoscop* OR "partial nephrectomy" OR endoscop* OR adrenalectomy OR cholecystectomy OR splenectomy OR nephrectomy OR "general surgery" OR "minimally invasive surgical procedur*" OR "minimally invasive surgery" OR "surgical procedur*" OR "operation room" OR " surgery robot*" OR "surgical robot*"

MeSH term: "Minimally Invasive Surgical Procedures"[Mesh]

"Surgical Procedures, Operative"[Mesh]

Search: "Minimally Invasive Surgical Procedures"[Mesh] OR "Surgical Procedures, Operative"[Mesh] OR laparoscop* OR "partial nephrectomy" OR endoscop* OR adrenalectomy OR cholecystectomy OR splenectomy OR nephrectomy OR "general surgery" OR "minimally invasive surgical procedur*" OR "minimally invasive surgery" OR "surgical procedur*" OR "operation room" OR " surgery robot*" OR "surgical robot*"

Concept 2: Deep learning

Keywords:"artificial intelligence" OR "machine learning" OR "Image-guided Surgery" OR "deep learning" OR "artificial neural network*" OR "neural network*" OR "convolutional neural network*"

MeSH terms: "Artificial Intelligence"[Mesh]

Search: "Artificial Intelligence"[Mesh] OR "artificial intelligence" OR "machine learning" OR "Image-guided Surgery" OR "deep learning" OR "artificial neural network*" OR "neural network*" OR "convolutional neural network*"

Concept 3: Video

Keywords: video* OR "Video Recording*" OR "Video-Assisted Surgery" MeSH term: "Video Recording"[Mesh]

Search: "Video Recording"[Mesh] OR video* OR "Video Recording*" OR "Video-Assisted Surgery"

### 1.4 Database Search Results

**Pubmed search: 20/12/2023**

("Minimally Invasive Surgical Procedures"[Mesh] OR "Surgical Procedures, Operative"[Mesh] OR laparoscop* OR "partial nephrectomy" OR endoscop* OR adrenalectomy OR cholecystectomy OR splenectomy OR nephrectomy OR "general surgery" OR "minimally invasive surgical procedur*" OR "minimally invasive surgery" OR "surgical procedur*" OR "operation room" OR " surgery robot*" OR "surgical robot*") AND ("Artificial Intelligence"[Mesh] OR "artificial intelligence" OR "machine learning" OR "Image-guided Surgery" OR "deep learning" OR "artificial neural network*" OR "neural network*" OR "convolutional neural network*") AND ("Video Recording"[Mesh] OR video* OR "Video Recording*" OR "Video-Assisted Surgery") = 2,288 results

**IEEE search: 17/12/2023**

(laparoscop* OR endoscop* OR surgery OR "minimally invasive surgery" OR "robot-assisted" OR “robot-assisted surgery” ) AND ( "artificial intelligence" OR "machine learning" OR "deep learning" OR “artificial neural network” OR “transformers” OR “Explainable AI” OR “Neural networks” OR “Data models” OR “convolutional neural network” OR “Task analysis”) AND ( video* OR "Video Recording*" OR "Video-Assisted Surgery" OR “Content based video retrieval” OR “indexing of surgical videos”) = 1,058 results

**Medline: 19/12/2023**

("Minimally Invasive Surgical Procedures" OR "Surgical Procedures" OR laparoscop* OR "partial nephrectomy" OR endoscop* OR adrenalectomy OR cholecystectomy OR splenectomy OR nephrectomy OR "general surgery" OR "minimally invasive surgical procedur*" OR "minimally invasive surgery" OR "surgical procedur*" OR "operation room" OR "surgery robot*" OR "surgical robot*") AND ( "artificial intelligence" OR "machine learning" OR "Image-guided Surgery" OR "deep learning" OR "artificial neural network*" OR "neural network*" OR "convolutional neural network*") AND (video* OR "Video Recording*" OR "Video-Assisted Surgery") = 1,208 results

**Embase: 17/12/2023**

(laparoscop* OR endoscop* OR 'surgery'/exp OR surgery OR 'minimally invasive surgery'/exp OR 'minimally invasive surgery' OR 'robot-assisted' OR 'robot-assisted surgery'/exp OR 'robot-assisted surgery') AND ('artificial intelligence'/exp OR 'artificial intelligence' OR 'machine learning'/exp OR 'machine learning' OR 'deep learning'/exp OR 'deep learning' OR 'artificial neural network'/exp OR 'artificial neural network' OR 'transformers' OR 'explainable ai' OR 'neural networks'/exp OR 'neural networks' OR 'data models' OR 'convolutional neural network'/exp OR 'convolutional neural network' OR 'task analysis') AND (video* OR 'video recording*' OR 'video-assisted surgery'/exp OR 'video-assisted surgery' OR 'content based video retrieval' OR 'indexing of surgical videos') = 2,522 results

**Scopus: 17/12/2023**

( TITLE-ABS-KEY ( laparoscop* OR endoscop* OR surgery OR "minimally invasive surgery" OR "robot-assisted" OR "robot-assisted surgery" ) AND TITLE-ABS-KEY ( "artificial intelligence" OR "machine learning" OR "deep learning" OR "artificial neural network" OR "transformers" OR "Explainable AI" OR "Neural networks" OR "Data models" OR "convolutional neural network" OR "Task analysis" ) AND TITLE-ABS-KEY ( video* OR "Video Recording*" OR "Video-Assisted Surgery" ) )= 2,148 results

**Web of Science: 19/12/2023**

(ALL=("Minimally Invasive Surgical Procedures") OR ALL=( "Surgical Procedures" ) OR ALL=(laparoscop*) OR ALL=("partial nephrectomy") OR ALL=(endoscop* ) OR ALL=(adrenalectomy) OR ALL=(cholecystectomy) OR ALL=(splenectomy ) OR ALL=("minimally invasive surgery") OR ALL=("surgical procedur*") OR ALL=("operation room") OR ALL=("surgery robot*") OR ALL=("surgical robot*")) AND ( ALL=( "artificial intelligence") OR ALL=( "Image-guided Surgery") OR ALL=( "machine learning") OR ALL=( "deep learning") OR ALL=( "artificial neural network*") OR ALL=( "neural network*") OR ALL=( "convolutional neural network*")) AND ( ALL=( "video* " ) OR ALL=("Video Recording*") OR ALL=("Video-Assisted Surgery")) = 1,248 results

## Appendix 2: Data extraction tables

Table 7: Overview of Studies on Deep Learning Algorithms for Surgical Instrument Annotation in Robotic-Assisted Surgeries

| Title | Purpose | Limitations | Annotation Type | Clinical Use Case | Deep Learning Algorithm | Code Link |
|---|---|---|---|---|---|---|
| Baek et al 2019 [1] | To integrate a hysteresis compensator with learning-based pose estimation for a flexible endoscopic surgery robot. | Limitations not provided | Binary segmentation | Not mentioned | Siamese Convolutional Neural Network (SCNN) | NA |
| Islam et al 2019 [2] | Accurate and efficient segmentation of the surgical scen and identification and tracking of instruments | Limitations not found | Binary, parts, instrument segmentation | Decision making | light-weight cascaded convolutional neural network with a multi-resolution feature fusion | https://github.com/mobarakol/Surgical_Instruments_Segmentation |
| Colleoni et al 2019 [3] | Propose a novel encoder–decoder architecture for surgical instrument joint detection and localization that uses three-dimensional convolutional layers to exploit spatio-temporal features from laparoscopic videos | Limited number of testing videos, which is due to the lack of available annotated data. 2D nature of the estimated joint position | Surgical-tool joint detection | -Training<br>- Skill assessment | encoder–decoder architecture for surgical instrument joint detection and localization that uses three-dimensional convolutional layers | NA |
| Suzuki et al 2019 [4] | To improve the accuracy of instrument segmentation on laparoscopic images captured by a monocular camera using depth estimation combined with color information. | Limitations not found | Instrument Segmentation | - Decision making<br>- Patient safety | Fully-convolutional neural network (FCN) for depth estimation and U-Net-based image segmentation method. | NA |

| | | | | | | |
|---|---|---|---|---|---|---|
| Hasan et al 2019 [5] | Modify U-Net architecture to improve the segmentation of surgical tools in laparoscopic videos. | Limitations not found | Semantic segmentation | Not mentioned | Modified U-Net architecture (U-NetPlus) | https://github.com/SMKamrulHasan/UNetPlus |
| Plishker 2019 [6] | Develop a framework for training dataset generation and segmentation of surgical instruments in laparoscopic videos. | Limitations not found | Instrument Segmentation | Not mentioned | Convolutional Neural Network (CNN) | NA |
| Cai et al 2020 [7] | New framework using two three-layer convolutional neural networks (CNNs) in series for detecting surgical instruments in in-vivo video frames. | Ineffectiveness in dealing with image blurring caused by the instrument's fast movement and restriction to straight instruments. | Edge detection of the instrument + Tip location of the instrument | Not mentioned | Two three-layer convolutional neural networks (CNNs) | NA |
| Ni et al 2020 [8] | To develop a lightweight network for real-time segmentation of robotic surgical instruments. | Limitations not found | Real-time segmentation of robotic surgical instruments | -prediction of dangerous operations<br>- skill assessment<br>-workflow optimization<br>-surgical report generation<br>-surgical task automation | attention-guided lightweight network (LWANet) | https://github.com/nizhenliang/LWANet |
| Lotfi 2020 [9] | detecting and tracking surgical instruments in vitreo-retinal eye surgery | may encounter difficulties with extremely blurred images or under conditions that significantly differ from the training data. | Skill Assessment and Training for Novice Surgeons, Surgical Analysis and Intervention, Integration with Surgical Systems | - training | | NA |

| | | | | | | |
|---|---|---|---|---|---|---|
| Kateryna Zinchenko And Kai-Tai Song 2021 [10] | To develop an autonomous endoscope system using instrument segmentation. | Limitations not found | real-time surgical instrument segmentation | -surgical navigation | YOLOv3 and ResNet Combined Neural Network (hybrid) | NA |
| Stoyanov et al 2021 [11] | Developing a method for surgical tool segmentation without manual annotations | Limitations not provided | Surgical tool segmentation | -surgical task automation<br>-skill assessment<br>-surgical workflow analysis | U-Net model | https://github.com/luiscarlosgph/semi-synthetic?tab=readme-ov-file |
| Colleoni 2021 [12] | Training deep learning models on datasets synthesized using image-to-image translation techniques for robotic instrument segmentation | Need for segmentation ground truth in producing real domain sets suitable for I2I model training | Synthetic frame generation for robotic instrument detection and segmentation | -surgical task automation<br>- pateint safety<br>- improve intra-operation | I2I model, encoder-decoder, U-Net | NA |
| Huang et al 2022 [13] | To segment surgical tools in endoscopic images using a pose-informed morphological polar transform. | Limitations not found | binary annotations of tool | - improve surgeon awareness | U-Net, Convolutional Neural Network | NA |

| | | | | | | |
|---|---|---|---|---|---|---|
| Huang et al 2022 [14] | To develop a fully unsupervised approach for surgical instrument segmentation. | Limitations not provided | binary Surgical Instrument Segmentation | -skill assessment | cycle-GAN | NA |
| Jin et al 2022 [15] | To explore intra- and inter-video relations for surgical semantic scene segmentation. | Limitations not provided | Surgical semantic scene segmentation | -surgical assessment<br>-surgical navigation<br>- surgical decision making | STswinCL | https://github.com/YuemingJin/STswinCL |
| Xu et al 2022 [16] | Developing a real-time, end-to-end surgical captioning model that eliminates the need for a heavy computational object detector or feature extractor. | The reliance on professional annotators for bounding box annotation and computational challenges in captioning models. | Surgical captioning tasks, particularly in the context of real-time robotic surgery. | Not mentioned | Shifted Window-Based Multi-Layer Perceptrons Transformer Captioning model (SwinMLP-TranCAP) | https://github.com/xumengyaamy/swinmlp_trancap |
| Tukra et al 2022 [17] | propose an unsupervised end-to-end deep learning framework, based on fully convolutional neural networks to reconstruct the view of the surgical scene under occlusions and provide the surgeon with intraoperative see-through vision in these areas | Limitations not found | segmenting instruments | - patient safety<br>- improve surgeon visual awareness | fully convolutional architecture based on a dual input stream encoder and a single decoder | NA |

| | | | | | | |
|---|---|---|---|---|---|---|
| Reiter 2022 [18] | propose the application of a transformer based architecture for end-to-end tool detection. | Limitations not found | tool detection | - surgery report generation<br>-decision making | R-CNN, DETR | NA |
| Leifman et al 2022 [19] | To propose a novel approach for instrument segmentation in laparoscopic surgeries that uses weak and fast annotations provided as bounding boxes of the instruments. The method uses synthetic images and CycleGAN for domain adaptation to enhance training data and improve segmentation accuracy. | Limitations not found | pixel-accurate segmentation of surgical instruments | -surgical workflow analysis<br>-patient safety<br>-skill assessment<br>-training<br>- postoperation outcomes analysis | cycle-GAN | NA |
| Zheng et al 2022 [20] | Automatic recognition of tools and episodes in robotic bronchoscopy using a multi-task architecture with Vision Transformers. | Difficulty in distinguishing between visually similar tools and occlusion scenarios. | Biopsy tool presence and episode recognition | -workflow analysis<br>- intra-operative monitoring to providing automated assistance to the clinical staff<br>-postoperation outcomes analysis | Multi-Task Vision Transformer (MT-ViT) architecture. | NA |
| Xia et al 2023 [21] | To develop a network for semantic segmentation of surgical instruments using a nested U-structure. | Limitations not found | Binary segmentation, Parts segmentation, type segmentation | - AR<br>- improve surgeon awareness<br>- skill assessment | Nested U-Structure Network combined with ResNet (hybrid) | NA |

| | | | | | | |
|---|---|---|---|---|---|---|
| Bian 2023 [22] | Estimate intra-operative motion under occlusion conditions | Limitations not found | Instrument Segmentation | Not mentioned | RAFT-GMA architecture & CNN-GRU | NA |
| Shubhangi Nema and Leena Vachhani 2023 [23] | Image segmentation of instruments in raw surgical videos for intraoperative assistance software using unpaired training | The requirement of specific labelled datasets for certain surgical tools and the method's dependency on such data | Multi-class segmentation of both robotic and rigid surgical instruments in raw videos | -skill assessment | InstruSegNet, an adversarial network | NA |
| Wang et al 2023 [24] | urgical report generation by explicitly exploring the interactive relation between tissues and surgical instruments. | Limitations not found | detection of surgical tools | -surgical report generation<br>- decision making<br>- postoperative outcome analysis<br>-patient safety | YOLOv5 | NA |
| Hayoz 2023 [25] | present a solution for stereo endoscopes that estimates depth and optical flow to minimize two geometric losses for camera pose estimation. | Limitations not found | surgical tool segmentation + localization | Not mentioned | encoder & decoder, U-Net, DDN | https://github.com/aimi-lab/robust-pose-estimator |

| | | | | | | |
|---|---|---|---|---|---|---|
| Brandenburg et al 2023 [26] | Extracting surgomic features from robot-assisted minimally invasive esophagectomy videos using active learning to reduce annotation effort. | Challenges in extracting surgomic features from highly variable surgical scenarios and data quality. | Multi classification of surgical tools | -postoperative outcomes analysis | ResNet18 | NA |
| Ping et al 2023 [27] | Developing a CNN and YOLOv3 based model for identifying surgical tools and tool tips in various surgical scenarios. | Variations in lighting, occlusion, and surgical scenes impacting tool recognition accuracy. | Tool and tool tip recognition | - training<br>-skill assessment | CNN and YOLOv3 algorithm | NA |
| Ayobi et al 2024 [28] | Instrument segmentation using transformers | Limitations not found | Surgical instrument segmentation | -pose estimation<br>- workflow analysis | Transformers with masked-attention mechanism | https://github.com/bcv-uniandes/matis |
| Law et al 2017 [29] | Assess technical skill level of surgeons by analyzing movement of robotic instruments in surgical videos, leveraging crowd workers for data, Hourglass Networks for instrument tracking, and a linear classifier for skill assessment. | Challenges include tracking surgical instruments due to occlusions and interactions between tissue and instruments, and reliance on video data quality. | Surgical-tool joint detection | -postoperative outcome analys<br>-skill assessment | Convolutional Neural Networks (ConvNets) with Hourglass Networks | NA |

| | | | | | | |
|---|---|---|---|---|---|---|
| Ross et al 2018 [30] | To investigate the concept of self-supervised learning for automatic annotation of medical images using unlabeled video data. | Limitations not provided | instrument segmentation | -improve postoperation outcomes | Conditional Generative Adversarial Network (cGAN) for auxiliary task (re-colorization); U-Net for target task (instrument segmentation) | NA |
| Jin et al 2019 [31] | Leverage instrument motion information for accurate segmentation by incorporating temporal prior to an attention pyramid network | Slightly higher standard deviations in cases of unexpected motion | surgical in-strument segmentation | -skill assessment<br>- postoperative outcome analysis | MF-TAPNet (Motion Flow - Temporal Attention Pyramid Network) | https://github.com/keyuncheng/MF-TAPNet |
| Kletz et al 2019 [32] | Segmentation and recognition of surgical instruments in videos recorded from laparoscopic gynecology. | High similarity of surgical instruments poses a challenge for accurate recognition. | segmentation and recognition of surgical instruments | -skill assessment | Region-based fully convolutional network | NA |
| Claudio S et al 2020 [33] | Tracking points on intra-operative retinal fundus videos through optical flow estimation | Performance might degrade due to less training data or simpler datasets | estimate opti- cal flow and provide a segmentation containing information about whether a pixel belongs to the retinal fundu | Not mentioned | convolutional neural network | NA |

| | | | | | | |
|---|---|---|---|---|---|---|
| Kalia et al 2021 [34] | Present a joint unpaired image-to-image mapping and segmentation strategy for better generalizability of a surgical instrument segmentation model to a domain with no labelled data. | Challenges in accurately translating domain-specific features and dealing with a lack of labeled data in target domains. | binary segmentation | Not mentioned | Generative Adversarial Networks (GANs), proposed model is coSegGAN | https://github.com/tajwarabraraleef/coSegGAN |
| Choi et al 2021 [35] | Implement robot vision for assisted surgery in simple mastoidectomy using a convolutional neural network. | Limitations not found | Detection and segmentation | -Training<br>-Skill assessment | U-Net, YOLACT, and YOLOv4 | NA |
| Hasan et al 2021 [36] | To perform surgical tool detection, segmentation, and 3D pose estimation in Computer-Assisted Laparoscopy (CAL). | Challenges in highly variable and complex laparoscopic imaging conditions. | tool Detection | Not mentioned | Convolutional Neural Networks (CNNs), Convolutional LSTM | https://github.com/kamruleee51/ART-Net |
| Yang et al 2022 [37] | Presenting Alternative method for annotating surgical instruments for semantic segmentation using weakly supervised learning. | The method may struggle with accurate ground truth annotations and could overfit to inaccurate training labels. | Surgical instrument semantic segmentation | Not mentioned | DeeplabV3plus10 with ResNet50 as encoder | NA |
| Kugener et al 2022 [38] | Assess the utility of the SOCAL dataset for machine learning applications in surgical video analysis. | Challenges in detecting surgical instruments in the presence of hemorrhage. | Instrument detection | -Postoperative outcome analysis<br>-Training<br>-Skill assessment | RetinaNet and YOLOv3 are convolutional neural network (CNN)–based deep neural networks (DNNs) | NA |

| | | | | | | |
|---|---|---|---|---|---|---|
| Wang et al 2023 [39] | To develop a deep learning algorithm for scoring GEARS and OSATS from robotic partial nephrectomy procedure videos. | Limited variability in the scores due to real surgeries; most surgeons are experts scoring highly. | segmenting instruments | -Skill assessment | Multi-task Convolutional Neural Network (mtCNN) | NA |
| Backer et al 2023 [40] | Enhance the safety of AR-guided surgery by real-time instrument detection during AR-guided robot-assisted partial nephrectomy and kidney transplantation | 0.5-s delay in real-time binary segmentation algorithm and manual alignment of 3D models | Instrument detection in augmented reality for surgical guidance | -AR | (EfficientNetB5) and (U-Net) | NA |
| Marullo et al 2023 [41] | Real-time blood accumulation detection and tools semantic segmentation from laparoscopic surgery video | Limitations not found | Semantic segmentation and event detection in laparoscopic surgery | -Postoperative outcome analysis | Multi-task Convolutional Neural Network (CNN) | NA |
| Wang et al 2023 [42] | Remove fog or smoke from endoscopic video to maintain a clear visual field in robot-assisted minimally invasive surgery. | Difficulty handling different levels of smoke density and various spatial domains in surgical scenes. | Surgical smoke removal | Not mentioned | Residual Swin transformer network. CNN | NA |
| Valderrama et al 2023 [43] | To provide a new experimental framework for holistic surgical scene understanding, introducing the PSI-AVA Dataset for various recognition tasks in robot-assisted radical prostatectomy videos. | Limitations not found | Recognition of various elements (phases, steps, instruments, actions) in surgical videos | Not mentioned | Deep Convolutional Neural Networks-based, R-CNN | https://github.com/bcv-uniandes/tapir |

| | | | | | | |
|---|---|---|---|---|---|---|
| . Yang et al 2023 [44] | To design and validate an automated method for evaluating technical proficiency in colorectal robotic surgery using artificial intelligence. | AI tools might struggle with interpreting videos of different formats from alternative recording systems. | tool detection and tracking | -Skill assessment | Mask R-CNN | NA |
| Li et al 2023 [45] | Address the issue of confusion regions in surgical scenes for instrument segmentation using a semi supervised framework. | Challenges with sparse annotated surgical videos, interferences like blood or illumination. | Instrument Segmentation | - Skill assessment<br>- Decision making | Spatial–temporal transformer and confusion-aware contrastive learning. | NA |
| Backer et al 2022 [46] | Annotation of robotic instruments in surgery to facilitate surgical artificial intelligence projects. | Requires significant time and expertise for annotation; variation in annotation times among different annotators. | Pixel Segmentation | Not mentioned | Multi-task convolutional neural network; CNN | NA |
| Lou et al 2022 [47] | Propose a semi-supervised segmentation network based on contrastive learning for segmenting medical images. | Limitations not found | Tools Segmentation | Not mentioned | Encoder, decoder, CNN | https://github.com/angeloucn/min_max_similarity |
| Du et al 2018 [48] | Present deep neural network for articulated multi-instrument 2-D pose estimation trained on detailed annotations of endoscopic and microscopic datasets. | Challenges in accurately detecting articulation in complex surgical environments and dealing with variable lighting and visibility conditions. | Fully Convolutional Networks | -Detecting the location of individual joint parts | Not mentioned | https://github.com/surgical-vision/EndoVisPoseAnnotation |

Table 8: Summary of Deep Learning Models and Their Technical Specifications for Surgical Instrument Annotation

| Title | Deep Learning Model | Network Architecture | Batch Size | Epochs | Learning Rate | Optimizer | Loss Function | Performance | Hardware |
|---|---|---|---|---|---|---|---|---|---|
| Baek et al 2019 [1] | Utilizes a Siamese Convolutional Neural Network (SCNN) | The SCNN consists of two branches with the same architecture and shared weights, designed for feature extraction and to determine the similarity between pairs of images. | Not mentioned | Not mentioned | Not mentioned | Not mentioned | Not mentioned | Not mentioned | NVIDIA RTX-2080ti GPU. |
| Islam et al 2019 [2] | Light weight cascaded convolutional neural network with a multi-resolution feature fusion | The architecture includes multiple branches for fusing contextual information at different resolutions, featuring Conv-Block, Residual-Block, Multi-resolution Feature Fusion (MFF) module, Decoder, and Up-sampling units. | Not mentioned | Not mentioned | 0.001 | Adam | Combination of main branch loss, auxiliary loss | Dice: 0.916 | 2 NVIDIA GTX 1080Ti GPUs |
| Colleoni et al 2019 [3] | Encoder–decoder architecture for surgical instrument joint detection | The 3D convolution allows the kernel to move along the three input dimensions to process multiple frames at the same time, preserving and processing temporal information through the network, using a modular encoder-decoder structure. We used a two-branch architecture to allow the FCNN to separately process the joint and connection mask | 2 | Not mentioned | 0.001 | SGD | Binary cross-entropy | Dice similarity coefficient of 85.1% for joint detection | Nvidia GeForce GTX 1080 |
| Suzuki et al 2019 [4] | Fully convolutional neural network (FCN) for depth estimation and U-Net-based image segmentation method. | U-Net is a kind of FCNs that contains two parts: an encoding part to extract image features and a decoding part to output a segmentation result. The U-Net has several skip connections to generate a high-resolution segmentation map. | Not mentioned | Not mentioned | Not mentioned | Not mentioned | Not mentioned | Binary seg -IoU: 66.9% - DiCE: 79.0 | V100 GPU |
| A. Linte 2019 [5] | Modified U-Net architecture (U-NetPlus) | Transfer learning for a new task is to partially reuse ImageNet feature extractor — VGG-11 or VGG-16 as encoder — and then add a decoder. | 4 | 100 | 0.00001. | Adam | Jaccard index | -Binary segmentation IoU:83.75 -Binary segmentation DiCE:90.20 - Instrument Part IoU:65.75 - Instrument part DiCE:76.26 | NVIDIA GTX 1080 Ti GPU (11GBs of memory). |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | - Instrument type IoU:34.19<br>- Instrument type DiCE: 45.32 | |
| Plishker 2019 [6] | Convolutional Neural Network (CNN) | The architecture comprises the convolutional encoder-decoder model LinkNet-34, where a pre-trained ResNet-34 network is used in the encoder, and refinements are made in the decoder parts. | 16 | 20 | Plishker 2019 [6] | Convolutional Neural Network (CNN) | Jaccard Index | DSC averaging 95.19% and Jaccard Index averaging 90.85% | Not mentioned |
| Cai et al 2020 [7] | Two three-layer convolutional neural networks (CNNs) | Two convolutional layers, two rectified linear units (ReLU), a fully connected (fc) layer, and a SoftMax layer, without pooling layers. | 125 for Convnet1, 100 for Convnet2 | 20 for Convnet1, 25 for Convnet2 | 10-4 for Convnet1, 10-3 for Convnet2 | SGD | Cross entropy | Accuracy of 75% on the EndoVis dataset | Not mentioned |
| Ni et al 2020 [8] | Attention-guided lightweight network (LWANet) | Encoder-decoder architecture with MobileNetV2 as the encoder. The decoder consists of depth wise separable convolution, attention fusion block, and transposed convolution. | 16 | Not mentioned | Initially 0.0002, adjusted by multiplying by 0.8 every 30 iterations | Adam | Focal loss | mIoU= 94.10% and mDice 96.91% | Nvidia Titan X GPU with 12G memory |
| Lotfi 2020 [9] | (YOLOv3) CNN | (YOLOv3) CNN is employed to detect the instruments. Thereafter, the Median- flow OpenCV tracker is utilized to track the determined objects. To modify the tracker, every” n” frames, the CNN runs over the image and the tracker is updated. Moreover, the dataset consists of 594 images in which four” shaft”,” center”, ”laser”, and ”gripper” labels are considered. | 400160 batches | Not mentioned | Not mentioned | Not mentioned | Not mentioned | Loss of 0.0812 | Nvidia GeForce GTX 1080 Ti GPU |
| Kateryna Zinchenko And Kai-Tai Song 2021 [10] | YOLOv3 and ResNet Combined Neural Network (hybrid) | YOLOv3 for instrument detection and ResNet for instrument segmentation, enhanced with morphological operations for real-time processing | Not mentioned | Not mentioned | Not mentioned | Not mentioned | Not mentioned | IoU of 86.6% | Implemented in ROS on Ubuntu, with visualization running on Windows in Unity3D |
| Stoyanov et al 2021 [11] | U-Net model trained on semi-synthetic data for surgical tool segmentation. | The architecture used is U-Net, which is a convolutional neural network designed for quick and precise segmentation tasks. | Not mentioned | Not mentioned | Not mentioned | Adam | Cross entropy | IoU:81.6 | NVIDIA Tesla V100 GPU (32 GB). |
| Colleoni 2021 [12] | I2I model, encoder-decoder, U-Net | The architecture involves two key components: generators and discriminators, forming a cycle-GAN framework. The generators | Not mentioned | 100 | 0.0002 | Adam | The model utilizes a combination of adversarial loss, cycle consistency | MICCAI 2020 real data IoU:96% MICCAI 2020 | NVIDIA Tesla V100 GPU |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | are responsible for translating images from the simulation domain to the real domain and vice versa, while the discriminators aim to distinguish between real and synthesized images. | | | | | loss, content loss, and structure similarity loss (SSIM) to train the I2I model effectively. The overall loss is a weighted sum of these individual loss components. | synthetic data IoU:92% | |
| Huang et al 2022 [13] | U-Net Convolutional Neural Network | U-Net enhanced with a pose-informed variable center morphological polar transform. The approach transforms tool shapes into more rectangular morphologies for improved segmentation. | 100 | 50 | 1 x 10^-4 | Adam | Dice Coefficient Loss | The network yielded near perfect algorithm selection and resulted in mean and median dice scores of 0.945 and 0.951 and mean and median IoU scores of 0.883 and 0.892. | Quad-core 64-bit Intel Core i7-7700, 16 GB DDR4 RAM, NVIDIA GeForce GTX 1070 with 8 GB GDDR5. |
| Huang et al 2022 [14] | cycle-GAN | Generative-adversarial network for unsupervised surgical tool segmentation of optical-flow images, generating pseudo-label masks. It then uses these masks to train a per-frame segmentation model through a learning-from-noisy-labels strategy. | 16 | 40&40 | $5 \times 10{-}5$ (÷2 / 5 epochs, after epoch 20) | Adam | Cross entropy | mIoU= 83.77 | Single NVIDIA Tesla V100 GPU (32 GB). |
| Jin et al 2022 [15] | STswinCL, a framework that integrates a hierarchy Transformer with a joint space-time window shift scheme for capturing intra-video relations | The architecture features a novel Transformer-based model with self-attention confined within local space-time windows. It employs a joint space-time window shift mechanism for efficient global context aggregation, alongside a segmentation head for pixel classification. | 8 | Not mentioned | 1e − 3 | SGD | Cross entropy | mIoU: 63.6% and dicce:72 | 2 NVIDIA RTX 3090 GPUs |
| Xu et al 2022 [16] | Shifted Window-Based Multi-Layer Perceptrons Transformer Captioning model (SwinMLP-TranCAP) | The SwinMLP-TranCAP model incorporates a vision encoder based on shifted window-based MLP and a Transformer-like decoder. It utilizes patch-based shifted window technique, replacing the multi-head attention module with window-based multi- | Not mentioned | 100 | 3e−4, | Adam | Cross entropy | SPI: 0.619 | NVIDIA RTX3090 GPU |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | head MLP for faster inference and reduced computation. | | | | | | | |
| Tukra et al 2022 [17] | Fully convolutional architecture based on a dual input stream encoder and a single decoder. | STV-Net features a densely connected encoder-decoder network combining 3D and 2D convolution operations to process both volumetric (sequence of video frames) and single image data. | Not mentioned | 3 epochs | 0.0002 | Adam | Combination of adversarial loss, reconstruction loss, perceptual loss, style loss, warping loss, and total variation loss. | SSIM:0.809 | NVIDIA P100 Tesla |
| Reiter 2022 [18] | R-CNN, DETR | DETR extracts image features with a backbone network, ResNet-50 pretrained on ImageNet, transformer-based architecture for end-to-end tool detection | 16 | 4 | 10−3, 10−4, 10− 5 | Adam | SVD | (0.66 mAP) | Not mentioned |
| Leifman et al 2022 [19] | cycle-GAN | CycleGAN | Not mentioned | Not mentioned | Not mentioned | Not mentioned | Not mentioned | - Bounding box & pixel accurate annotation DiCE: 0.941 | Not mentioned |
| Zheng et al 2022 [20] | Multi-Task Vision Transformer (MT-ViT) architecture. | Feature extractor: ResNet34, a 34-layer residual convolutional network, is used as shared backbone between frame- level and episode-level tasks to extract visual feature vectors from frames. Vision Transformer Encoder: Inspired by ViT, we utilize a Transformer encoder for the episode recognition task. The Transformer encoder can collect spatial- temporal information across the frames within each episode to optimize predictions by using its global receptive field and self-attention properties. | 24 | 50 | 5e−5 | Adam | Cross-entropy | Tools classification - accuracy: 91.53±0.13 - precision: 86.62±0.32 - recall: 87.07±0.21 | NVIDIA Quadro P6000 24GB GPU |
| Xia et al 2023 [21] | Nested U-Structure Network combined with ResNet (hybrid) | The network is a two-level nested U-structure based on the architecture of UNet. Each layer uses a U-structure rather than simple superposition of convolutional layers. It employs ResUNet as the main encoder, with ResUNet34 and ResUNetpp modules, along with RSU-4F at the bottom to maintain resolution. Hypbrid (resnet combined with u-net) | 2 | 100 | 1.00E-04 | Adam | The segmentation loss function is based on the Jaccard index. For binary segmentation, BCEWithLogitsLoss is used. For multi-class segmentation, Cross-Entropy Loss is used. | -Binary segmentaion IoU:83.75 - binary segmentation DiCE:90.20 - Instrument Part IoU:65.75 - Instrument part DiCE:76.26 - Instrument type IoU:34.19 | 2 NVIDIA GTX3090 GPUs |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | - Instrument type DiCE: 45.32 | |
| Bian 2023 [22] | RAFT-GMA architecture & CNN-GRU | The Motion Decoupling Network architecture integrates a segmentation subnet that estimates the segmentation map of instruments in an unsupervised manner. It employs separate constraints for tissue and instrument motion and utilizes a hybrid self-supervised strategy with occlusion completion to enhance learning efficacy. | 8 | 80 | 0.0001 | Adam | A combination of photometric loss, smoothness loss, consistency loss, and segmentation loss | Not mentioned | NVIDIA TITAN Xp GPU |
| Shubhangi Nema and Leena Vachhani 2023 [23] | InstruSegNet, an adversarial network optimized on least square Generative Adversarial Networks loss | The network combines features from both U-Net and ResNet, introducing an enhanced generator network architecture. It consists of an encoding and decoding path with multiple identity blocks in between. | Not mentioned | 100 | 0.0002 | Adam | Least Squares Generative Adversarial Networks (LSGAN) l | Robotic Instrument Segmentation: Average pixel accuracy of 97.31% for single instruments and 92.32% for multiple instruments; specificity of 99.142% for single instruments and 97.806% for multiple instruments. Rigid Instrument Segmentation: Average pixel accuracy of 91.4% and specificity of 94.94%. | i7 processor, high-performance NVIDIA processor, and 16 GB RAM. |
| Wang et al 2023 [24] | YOLOv5 | The architecture includes a region feature extraction part, a relational exploration module, an interaction perception module, and a caption generation part. It employs YOLOv5 for object detection, ResNet18 for feature extraction, a node tracking mechanism, and the | 50 | 80 | 0.00006 | Adam | BCEWithLogits loss and CIoU loss for object detection | BLEU-1 (11.97%), BLEU-2 (12.94%), BLEU-3 (13.48%), BLEU-4 (12.96%), ROUGE | NVIDIA GeForce RTX 2080 Ti GPU with 11 GB memory |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | M2 transformer for surgical report generation. | | | | | | (8.36%), and CIDEr (36.63%) | |
| Hayoz 2023 [25] | Encoder & decoder, U-Net, DDN | DeepLabv3+ architecture | 8 | 200 | 10−5. | Adam | Not mentioned | Not mentioned | NVIDIA RTX3090 GPU |
| Brandenburg et al 2023 [26] | ResNet18 | Multiple Bayesian ResNet18 architectures were trained on a multicentric dataset | 16 | 100 | 3e–3 | SGD | Not mentioned | Not mentioned | Nvidia RTX A5000 NVIDIA Corporation, (Santa Clara, California, USA). |
| Ping et al 2023 [27] | CNN and YOLOv3 algorithm | Backbone Network (BN), feature enhancement, and detection head, DarkNet-19 was used as BN. Each prediction result included three parts: confidence, category, and coordinates. Only when the confidence was greater than a certain threshold (0.8 in our present work) could it be considered as the detection result output by the model. Detection boxes were generated after post-processing. | Not mentioned | 100 | 0.002. | SGD | Not mentioned | - Endoscopic surgical tool tips recognition precision: 94.53% (13,278/14,045) with an IoU threshold of 0.1 | NVIDIA Tesla V100 |
| Ayobi et al 2024 [28] | Transformers with masked-attention mechanism | The architecture comprises a masked attention transformer baseline for creating and classifying region proposals. An independent temporal consistency module is employed to leverage video-level information for improved classification. Mask2Former with a Swin Transformer backbone is used as the baseline to leverage multi-scale deformable attention and masked attention mechanisms. | 24 | 100 | Not mentioned | ADAMW | Not mentioned | Endovis17 mIoU: 71.36 ±3.46 Endovis18: 84.26 | 4 NVIDIA Quadro RTX 8000 GPUs for the masked attention baseline and a single NVIDIA Quadro RTX 8000 GPU |
| Law et al 2017 [29] | Convolutional Neural Networks (ConvNets) with Hourglass Networks for instrument tracking and linear classifier for skill assessment. | stack two hourglass modules in our network | 64 | Not mentioned | 0.001 | Not mentioned | Not mentioned | Not mentioned | Not mentioned |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Ross et al 2018 [30] | Conditional Generative Adversarial Network (cGAN) for auxiliary task (re-colorization); U-Net for target task (instrument segmentation) | CGAN-based architecture for re-colorization with U-Net architecture adapted for the segmentation task. The U-Net includes batch normalization and tanh re-scaling in the output layer. | 12 | 150 | 0.0005 for the generator and 0.002 for the discriminator in the cGAN; 0.0005 for the segmentation task | Adam | Cross-entropy loss for segmentation | DSC for dataset fraction: 1/8 mean:0.68 (10%) | Not mentioned |
| Jin et al 2019 [31] | MF-TAPNet (Motion Flow - Temporal Attention Pyramid Network) | The architecture includes an encoder-decoder segmentation network with an attention pyramid. It utilizes motion flow to propagate the prediction mask of the previous frame to the current frame, deriving a temporal prior that is incorporated at the bottleneck layer of the segmentation network as an initial attention map. | 8 | Not mentioned | Initialized as 3e−5 for binary and part segmentation tasks, and 2e−5 for the type of segmentation task. | Adam | cross entropy | IoU (Intersection over Union) of 87.56% and a Dice score of 93.37%. | 4 NVIDIA Titan GPUs. |
| Kletz et al 2019 [32] | Region-based fully convolutional network | Mask R-CNN with ResNet-101 as backbone | Not mentioned | Not mentioned | Not mentioned | Not mentioned | Binary cross-entropy | - Segmentation average precision: 81% | Not mentioned |
| Claudio S et al 2020 [33] | Convolutional neural network | The architecture utilizes the "FlowNetSimple" architecture, which is a fully convolutional network with an encoder-decoder structure and skip connections. | 10 | 100 | 10^-4 | Adam | The cost function consists of four loss terms: flow loss, regularization loss (to reduce overfitting), segmentation loss, and a total variation loss (to promote a smooth flow field). | Grid EPE:15.8 (7.9) | NVIDIA Quadro P6000 GPU, an Intel i7-6900K 16-core CPU, and 64 GB RAM. |
| Kalia et al 2021 [34] | Generative Adversarial Networks (GANs), proposed model is coSegGAN | The generative part of coSegGAN uses a cycleGAN-like architecture with two generators and two discriminators. The segmentation model in coSegGAN uses the original U-Net architecture with 16 base filters. | 8 | 100 | 10 −3 | Adam | Combination of a focal loss variant for segmentation, adversarial loss, pixel-level cycle consistency loss | Δ Dice = 0.9% | NVIDIA Tesla V100 GPU (16GB) |
| Choi et al 2021 [35] | U-Net, YOLACT, and YOLOv4 | U-Net, YOLOv4, and YOLACT | Not mentioned | Not mentioned | Not mentioned | Not mentioned | Cross-entropy | mAP of Tool detection: 94.7 | TITAN RTX environment. |
| Hasan et al 2021 [36] | CNN | Convolutional Neural Networks (CNNs), Convolutional LSTM, t has a Single Input Multiple Output | Not mentioned | Not mentioned | 1.0 (initial) | Adadelta | Binary or categorical cross-entropy | -Tool deetction accuracy: 100% -Tool detection | Not mentioned |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | (SIMO) architecture with one encoder and multi- ple decoders to achieve detection, segmentation, and geometric primitive extraction | | | | | | average precision: 100% -Tool segmentation mIoU: 81% -Tool segmentation accuracy:100% | |
| Yang et al 2022 [37] | CNN | DeeplabV3plus10 with ResNet50 as encoder | 8 | Not mentioned | 10−3 | Adam | Binary cross-entropy | 86.82% Intersection over Union (IoU) and 85.70% Dice | Not mentioned |
| Kugener et al 2022 [38] | Deep Neural Networks (DNNs) | RetinaNet and YOLOv3 are convolutional neural network (CNN)–based deep neural networks (DNNs) developed to detect spatial patterns within imagesYOLOv3 is a single-stage object detection method that uses CNNs to predict bounding boxes and assign probabilities to class types from the entire image. | Not mentioned | Not mentioned | Not mentioned | Not mentioned | Not mentioned | mAP = 0.67 of Instrument Detection | Not mentioned |
| Wang et al 2023 [39] | Multi-task Convolutional Neural Network (mtCNN) | The network uses mechanisms for weighting different portions of the video via an automated process known as attention in neural networks. Three different attention mechanisms were investigated: Weight Gated, Self-Attention, and Direct Self-Attention Pooling | 150,000 | 2,800 epochs | Not mentioned | Adam | Cross entropy | The training data IoU converge to about 0.8 and the evaluation data converge to an IoU of 0.65, | Wang et al 2023 [39] |
| De Backer et al 2023 [40] | Deep learning convolutional neural network with U-Net infrastructure | The architecture comprises an encoder (EfficientNetB5) that extracts meaningful information for instrument segmentation, followed by a decoder (U-Net) that projects this information onto the original surgical image. | Not mentioned | Not mentioned | Not mentioned | Not mentioned | Not mentioned | IoU score of 94.4% and a Dice score of 97.10% | Nvidia RTX3090 |
| Marullo et al 2023 [41] | Multi-task Convolutional Neural Network (CNN) | U-Net architecture with modifications. The backbone of the network is based on the encoder of U-Net, which has been slightly modified. The network includes a shared global feature extractor followed by two separate output branches for semantic segmentation and event detection. | 32 | 30 | 0.0001 | Adam | Binary Cross Entropy | Semantic segmentation: Dice Score of 81.89% | NVIDIA Quadro P4000 GPU |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Wang et al 2023 [42] | Residual Swin transformer network. CNN | Feature Extraction: Utilizes convolutional layers to capture shallow features.<br>Further Feature Extraction: Employs Swin Transformer Blocks for deep feature extraction, followed by a convolutional layer. | 8 | Not mentioned | 10^-4. | Adam | Perceptual Loss | SSIM:0.8567 | Two NVIDIA 3080 GPUs. |
| Valderrama et al 2023 [43] | Deep Convolutional Neural Networks-based, R-CNN | TAPIR leverages Vision Transformers and builds upon the Multiscale Vision Transformer (MViT) model. | 24 | 100 | 0.0125 | SGD | Binary cross-entropy | IoU:80.85 ± 1.54 | Quadro RTX 8000 GPU |
| Yang et al 2023 [44] | Mask R-CNN | The study utilized Mask R-CNN for tool instance segmentation, with a ResNet and feature pyramid network (FPN) backbone for feature extraction, and a region proposal network (RPN) for object bounding box proposals. | Not mentioned | Not mentioned | Not mentioned | Not mentioned | Not mentioned | (mAP) of 0.71 | Not mentioned |
| Li et al 2023 [45] | Semisupervised framework with a spatial–temporal transformer and confusion-aware contrastive learning. | The architecture employs a spatial–temporal transformer (STTransformer) equipped with a 3-D relative distance regression (RDR) mechanism for structural relation modeling, and a confusion-aware contrastive learning strategy for pixel relation modeling. | 8 | 150 | 1.00E-04 | Adam | Supervised loss for labeled data, pseudo supervised loss for unlabeled data, | mIoU:68.35 mDiCE:77.54 | NVIDIA RTX A6000 GPU. |
| De Backer et al 2022 [46] | Multi-task convolutional neural network; CN | Modified U-Net architecture, 2 × 2 max pooling operation with stride 2, to halve the x-y image size. | 32 | 30 | 0.0001 | Adam | Cross Entropy Loss function | Dice Score equal to 81.89% for the semantic segmentation | NVIDIA Quadro P4000 GPU |
| Lou et al 2022 [47] | Encoder, decoder, CNN | Res2Net as an encoder which is pre-trained on ImageNet. The decoder contains four stages, and each stage contains a convolution block and an up-sampling layer, resulting in a predicted segmentation mask at the original image resolution. | Not mentioned | Not mentioned | Not mentioned | Not mentioned | Not mentioned | Tool segmentation<br>- DSC: 0.921<br>- IoU:0.921 | RTX A5000 GPU, CUDA 11.3 and cuDNN V8.2.1 |
| Du et al 2018 [48] | fully convolution network | FCN model has a bi-branch architecture with an encoder-decoder prediction process. It includes joint detection and association subnetworks, followed by a regression subnetwork. | Not mentioned | 10 | Initially, 0.001 with a decrease every 10 epochs by 5% | SGD | Cross entropy | Recall score of 82.99% and a precision score of 83.70%. | single Nvidia GeForce GTX Titan X GPU |

Appendix 3

# PRISMA 2020 Checklist

| Section and Topic | Item # | Checklist item | Location where item is reported |
|---|---|---|---|
| **TITLE** | | | |
| Title | 1 | Identify the report as a systematic review. | Page 1 |
| **ABSTRACT** | | | |
| Abstract | 2 | See the PRISMA 2020 for Abstracts checklist. | DNR |
| **INTRODUCTION** | | | |
| Rationale | 3 | Describe the rationale for the review in the context of existing knowledge. | Pages 3-4 |
| Objectives | 4 | Provide an explicit statement of the objective(s) or question(s) the review addresses. | Page 4 |
| **METHODS** | | | |
| Eligibility criteria | 5 | Specify the inclusion and exclusion criteria for the review and how studies were grouped for the syntheses. | Pages 5-6 |
| Information sources | 6 | Specify all databases, registers, websites, organisations, reference lists and other sources searched or consulted to identify studies. Specify the date when each source was last searched or consulted. | Page 5 |
| Search strategy | 7 | Present the full search strategies for all databases, registers and websites, including any filters and limits used. | Page 5 |
| Selection process | 8 | Specify the methods used to decide whether a study met the inclusion criteria of the review, including how many reviewers screened each record and each report retrieved, whether they worked independently, and if applicable, details of automation tools used in the process. | Page 5 |
| Data collection process | 9 | Specify the methods used to collect data from reports, including how many reviewers collected data from each report, whether they worked independently, any processes for obtaining or confirming data from study investigators, and if applicable, details of automation tools used in the process. | Page 6 |
| Data items | 10a | List and define all outcomes for which data were sought. Specify whether all results that were compatible with each outcome domain in each study were sought (e.g. for all measures, time points, analyses), and if not, the methods used to decide which results to collect. | Page 5 |
| | 10b | List and define all other variables for which data were sought (e.g. participant and intervention characteristics, funding sources). Describe any assumptions made about any missing or unclear information. | Page 5 |
| Study risk of bias assessment | 11 | Specify the methods used to assess risk of bias in the included studies, including details of the tool(s) used, how many reviewers assessed each study and whether they worked independently, and if applicable, details of automation tools used in the process. | Not Applicable |
| Effect measures | 12 | Specify for each outcome the effect measure(s) (e.g. risk ratio, mean difference) used in the synthesis or presentation of results. | Not Applicable |
| Synthesis methods | 13a | Describe the processes used to decide which studies were eligible for each synthesis (e.g. tabulating the study intervention characteristics and comparing against the planned groups for each synthesis (item #5)). | Page 5 |
| | 13b | Describe any methods required to prepare the data for presentation or synthesis, such as handling of missing summary statistics, or data conversions. | Not Applicable |
| | 13c | Describe any methods used to tabulate or visually display results of individual studies and syntheses. | Not Applicable |
| | 13d | Describe any methods used to synthesize results and provide a rationale for the choice(s). If meta-analysis was performed, describe the model(s), method(s) to identify the presence and extent of statistical heterogeneity, and software package(s) used. | Page 6 |
| | 13e | Describe any methods used to explore possible causes of heterogeneity among study results (e.g. subgroup analysis, meta-regression). | Not Applicable |
| | 13f | Describe any sensitivity analyses conducted to assess robustness of the synthesized results. | Not Applicable |
| Reporting bias assessment | 14 | Describe any methods used to assess risk of bias due to missing results in a synthesis (arising from reporting biases). | Not Applicable |
| Certainty assessment | 15 | Describe any methods used to assess certainty (or confidence) in the body of evidence for an outcome. | Not Applicable |

## PRISMA 2020 Checklist

| Section and Topic | Item # | Checklist item | Location where item is reported |
|---|---|---|---|
| **RESULTS** | | | |
| Study selection | 16a | Describe the results of the search and selection process, from the number of records identified in the search to the number of studies included in the review, ideally using a flow diagram. | Page 6 |
| | 16b | Cite studies that might appear to meet the inclusion criteria, but which were excluded, and explain why they were excluded. | Not Applicable |
| Study characteristics | 17 | Cite each included study and present its characteristics. | Page 6-12 |
| Risk of bias in studies | 18 | Present assessments of risk of bias for each included study. | Not Applicable |
| Results of individual studies | 19 | For all outcomes, present, for each study: (a) summary statistics for each group (where appropriate) and (b) an effect estimate and its precision (e.g. confidence/credible interval), ideally using structured tables or plots. | Page 6-12 |
| Results of syntheses | 20a | For each synthesis, briefly summarise the characteristics and risk of bias among contributing studies. | Not Applicable |
| | 20b | Present results of all statistical syntheses conducted. If meta-analysis was done, present for each the summary estimate and its precision (e.g. confidence/credible interval) and measures of statistical heterogeneity. If comparing groups, describe the direction of the effect. | Page 6-12 |
| | 20c | Present results of all investigations of possible causes of heterogeneity among study results. | Page 6-12 |
| | 20d | Present results of all sensitivity analyses conducted to assess the robustness of the synthesized results. | Page 6-12 |
| Reporting biases | 21 | Present assessments of risk of bias due to missing results (arising from reporting biases) for each synthesis assessed. | Not Applicable |
| Certainty of evidence | 22 | Present assessments of certainty (or confidence) in the body of evidence for each outcome assessed. | Not Applicable |
| **DISCUSSION** | | | |
| Discussion | 23a | Provide a general interpretation of the results in the context of other evidence. | Page 12 |
| | 23b | Discuss any limitations of the evidence included in the review. | Page 15 |
| | 23c | Discuss any limitations of the review processes used. | Page 15 |
| | 23d | Discuss implications of the results for practice, policy, and future research. | Page 16 |
| **OTHER INFORMATION** | | | |
| Registration and protocol | 24a | Provide registration information for the review, including register name and registration number, or state that the review was not registered. | NA |
| | 24b | Indicate where the review protocol can be accessed, or state that a protocol was not prepared. | NA |
| | 24c | Describe and explain any amendments to information provided at registration or in the protocol. | NA |
| Support | 25 | Describe sources of financial or non-financial support for the review, and the role of the funders or sponsors in the review. | Title Page |
| Competing interests | 26 | Declare any competing interests of review authors. | Title Page |
| Availability of data, code and other materials | 27 | Report which of the following are publicly available and where they can be found: template data collection forms; data extracted from included studies; data used for all analyses; analytic code; any other materials used in the review. | Pages 8-9 |

*From:* Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021;372:n71. doi: 10.1136/bmj.n71