Medical Image Analysis 77 (2022) 102365

Listas de contenidos disponibles enCienciaDirectat

Análisis de imágenes médicas

jrevista Página de inicio:www.elsevier.com/locate/media

Navegación quirúrgica sin marcadores de la flexión de la varilla utilizando un estéreo neural red y realidad aumentada en fusión espinal

Marco von Atzigen [a, b, *], Florentin Liebmann [a, b], Armando Hoch [a, c], José Miguel Spirig [c], Mazda Farshad [c], Jess Snedeker [b, c], Philipp Fürnstahl [a]

[a] a Investigación en Informática Ortopédica, Hospital Universitario Balgrist, Universidad de Zúrich, Zúrich, Suiza
b Laboratorio de Biomecánica Ortopédica, ETH Zurich, Zurich, Suiza
C Departamento de Ortopedia, Hospital Universitario Balgrist, Universidad de Zúrich, Zúrich, Suiza

INFORMACION DEL ARTICULO

Historial del artículo:
Recibido el 22 de junio de 2021
Revisado el 16 de noviembre de 2021
Aceptado el 10 de enero de 2022
Disponible en línea el 22 de enero de 2022

Palabras clave:
Red neuronal estéreo
Realidad aumentada
Flexión de varillas
navegación quirúrgica

RESUMEN

La instrumentación de las cirugías de fusión espinal incluye la colocación de tornillos pediculares y la implantación de varillas. Si bien se han propuesto varios enfoques de navegación quirúrgica para la colocación de tornillos pediculares, se ha prestado menos atención a la guía de la adaptación específica del paciente del implante de varilla. Proponemos un enfoque de Realidad Aumentada (AR) intuitivo y sin marcadores para navegar por el proceso de flexión requerido para la implantación de varillas. Se entrena una red neuronal estéreo a partir de las transmisiones de video estéreo de Microsoft HoloLens de un extremo a otro para determinar la ubicación de las cabezas de los tornillos pediculares correspondientes. A partir de las posiciones digitalizadas de la cabeza del tornillo, se calcula la forma óptima de la varilla, traducida a un conjunto de curvaturas
parámetros, y se utiliza para guiar al cirujano con un nuevo enfoque de navegación. En la navegación basada en AR, el cirujano es guiado paso a paso en el uso de las herramientas quirúrgicas para lograr un resultado óptimo. Hemos evaluado el rendimiento de nuestro método en cadáveres humanos frente a dos métodos de referencia, a saber, la flexión convencional a mano alzada y la navegación de flexión basada en marcadores en términos de tiempo de flexión.
y maniobras de reflexión. Conseguimos un tiempo medio de plegado de 231 s con 0,6 maniobras de plegado por barra en comparación con los 476 s (3,5 plegados) y 348 s (1,1 plegados) obtenidos con nuestro freehand y
puntos de referencia basados en marcadores, respectivamente.

1. INTRODUCCION

La cirugía de fusión espinal está indicada para una variedad de trastornos espinales que incluyen deformidad, trauma, enfermedad degenerativa del disco, escoliosis y espondilolistesis (Martin et al., 2019). En el transcurso de la cirugía, los implantes de tornillo se insertan bilateralmente en los pedículos de las vértebras patológicas y se fusionan con un implante de varilla para formar una conexión rígida. Con una tasa de complicaciones de hasta el 15 % (Nasser et al., 2010; Barbanti-Brodano et al., 2020), el tratamiento quirúrgico de la columna sigue siendo un gran desafío porque el cirujano tiene que operar cerca de estructuras anatómicas vitales como el médula espinal, raíces nerviosas y arterias.

Debido al alto riesgo de lesiones, la cirugía de columna fue una de las primeras disciplinas quirúrgicas que aprovechó la navegación quirúrgica para permitir más ejecución quirúrgica más precisa y segura (Merloz et al., 1998; Mavrogenis et al., 2013). La colocación de tornillos pediculares es la más frecuente.
paso quirúrgico navegado en cirugía espinal (Laine et al., 1997; Schlenzka et al., 20 0 0; Merloz et al., 1998; Richter et al., 2005; Nottmeier y Crosby, 2007; Liebmann et al., 2019), mientras que la navegación del proceso de doblado de varillas sigue siendo un campo casi inexplorado. La navegación de la colocación de tornillos pediculares se basa en marcadores rastreados externamente en combinación con técnicas de imágenes médicas como Tomografía computarizada (TC) o fluoroscopia para hacer coincidir un plan de intervención generado preoperatoriamente con la anatomía intraoperatoria. Después de un registro exitoso, se puede navegar por el plan al proporcionar al cirujano los puntos de entrada de tornillos deseados y las trayectorias de perforación.

Después de la instrumentación con tornillos pediculares, se debe adaptar un implante de barra a la anatomía del paciente para que encaje en la forma de U. apertura de las cabezas sueltas de los tornillos pediculares (ver Fig. 1 c). El banco de doblado quirúrgico ilustrado en la Fig. 1a se usa para doblar la barra

* Autor correspondiente.
E-mail address: marco.vonatzigen@balgrist.ch (M. von Atzigen).

https://doi.org/10.1016/j.media.2022.102365

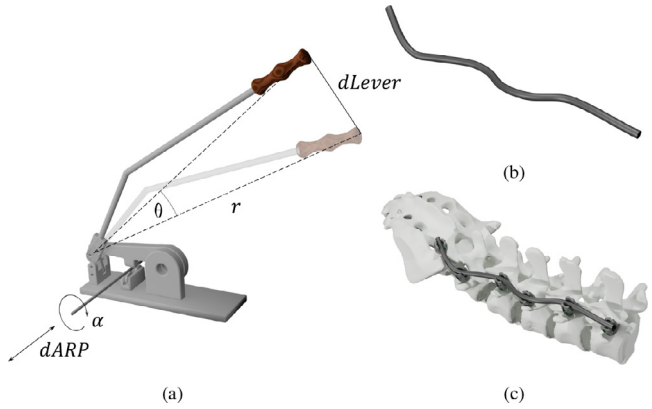


Fig. 1. (a) El banco de flexión que utilizan los cirujanos intraoperatoriamente para llevar la varilla a la forma deseada. La distancia que se debe mover la varilla para ajustar la posición axial de la varilla se denota por d ARP, la rotación axial requerida se da por α y el desplazamiento de la punta de la palanca se denomina d Le v er. (b) El implante de varilla doblada antes de la inserción en la anatomía instrumentada. (c) La situación deseada en la que la barra encaja perfectamente en las aberturas en forma de U de las cabezas de los tornillos pediculares colocados anteriormente.

implante intraoperatorio. Para cada intento de flexión, la barra axial
La posición (ARP) en el banco de doblado debe ajustarse para garantizar
flexión antee en la ubicación deseada. Además, el implante debe girarse axialmente durante la flexión para permitir curvas 3D, ya que el banco de flexión aplica la fuerza de flexión siempre en una dirección constante. Eventualmente, la flexión se ejecuta presionando hacia abajo la palanca del banco de flexión (ver Fig. 1 a). En la práctica clínica, los cirujanos eligen cada uno de los parámetros antes mencionados por juicio visual y ejecutan la flexión en función de su experiencia. Cada implante de varilla requiere varios pasos de flexión hasta que se alcanza su forma final, como sugiere la Fig. 1 (b). En el curso del proceso de doblado, los cirujanos tienen que moverse hacia adelante y hacia atrás entre el banco de doblado y la anatomía para refinar iterativamente la forma de la barra. Las adaptaciones de la forma de la barra después del inicio del proceso de implantación se denominan maniobras de reflexión y se asocian con inferior
calidad de varilla y tiempo quirúrgico prolongado. Por lo tanto, la reducción del número de intentos de reflexión es un requisito clínico y un objetivo quirúrgico. El procedimiento finaliza cuando el implante de varilla encaja suavemente en las cabezas de los tornillos pediculares en forma de U. La desalineación del implante de barra con respecto a las cabezas de los tornillos pediculares provoca una reducción forzada de la barra que, a su vez, puede provocar la pérdida del tornillo.
aflojamiento o incluso extracción (Wanivenhaus et al., 2019). El COM-
La combinación de desalineación de varillas y esterilidad comprometida dependiente del tiempo (Menekse et al., 2015; Uzun et al., 2020; Dalstrom et al., 2008) que surge del movimiento de ida y vuelta entre el banco de flexión y la anatomía promueve el uso de navegación quirúrgica de este paso quirúrgico crucial.

Hasta ahora, solo algunos enfoques para la navegación de la flexión de la espina.l

Se han propuesto implantes de varilla. Un producto médico comercial que aborda el proceso de doblado es Bendini (NuVasive, Inc., San Diego, CA, EE. UU.), que inicialmente requiere que los cirujanos capturen las ubicaciones de los tornillos pediculares implantados utilizando un dispositivo de puntería con seguimiento óptico. Las ubicaciones de las cabezas de los tornillos pediculares obtenidas se traducen luego a un modelo informático de la forma de varilla deseada. En un paso siguiente, la forma de la varilla objetivo se convierte en un conjunto de posiciones de flexión y ángulos de flexión que se proporcionan a los cirujanos para que los ejecuten con una herramienta de flexión patentada. Un estudio
demostraron que la navegación del proceso de doblado con su solución redujo las fuerzas ejercidas sobre los tornillos durante la implantación de la barra (Tohmeh et al., 2014). A pesar de que su estudio respalda el uso de métodos asistidos por computadora para el proceso de doblado de varillas, el sistema nunca se ha convertido en lo último en tecnología. En opinión de los autores, esto puede explicarse por la preparación adicional y
esfuerzo de mantenimiento, las herramientas quirúrgicas propietarias y el hardware costoso, como un sistema de seguimiento externo, que necesita
ser instalado y calibrado en la sala de operaciones (OR) para cada intervención.
Un enfoque más reciente de nuestro grupo de investigación explota la Realidad Aumentada (AR) para la navegación quirúrgica del proceso de doblado de barras (Wanivenhaus et al., 2019). Los tornillos pediculares implantados se digitalizan utilizando un dispositivo señalador esterilizado equipado con un marcador que es rastreado por las cámaras del HoloLens 1 montado en la cabeza (Microsoft, Redmond, WA, EE. UU.). Basándose en las posiciones de la cabeza del tornillo digitalizadas manualmente, la forma deseada de la barra se calcula mediante un spline centrípeto de Catmull-Rom (Barry y Goldman, 1988)
y presentado al cirujano como un holograma tridimensional (3D). Aunque este estudio pudo demostrar una reducción estadísticamente significativa del tiempo de flexión y los intentos de flexión, la tarea de flexión en sí siguió siendo un desafío, especialmente para las deformidades complejas. Aparentemente, el enfoque de simplemente presentar el deseado
la forma de la barra proporcionó una guía insuficiente para el cirujano. Además, el dispositivo señalador entra en contacto con el cirujano y el paciente, lo que complica significativamente el flujo de trabajo quirúrgico debido a los pasos adicionales de esterilización y calibración.
En este estudio, abordamos los inconvenientes antes mencionados al combinar un método puramente basado en la visión para estimar las posiciones de la cabeza del tornillo con instrucciones paso a paso novedosas basadas en AR para lograr la navegación de flexión. Se logra una estimación robusta y precisa de las posiciones 3D de las cabezas de los tornillos al extender nuestra anterior detección de cabezas de tornillos de flujo único (von Atzigen et al., 2020) a una red neuronal estéreo. Esta arquitectura detecta y asocia
cabezas de tornillo correspondientes de un extremo a otro en imágenes estéreo capturadas por las cámaras ambientales izquierda y derecha del
HoloLens 1. Otra contribución importante es la navegación basada en AR propuesta que guía el manejo preciso del banco de flexión superponiendo hologramas.
Nuestro método propuesto se validó con nuestro método AR anterior basado en marcadores (Wanivenhaus et al., 2019), así como con el proceso a mano alzada de última generación en términos de tiempo de flexión y el número de maniobras de flexión en cadáveres humanos.

2. Trabajo relacionado

La sección de trabajos relacionados se estructura en dos partes. Primero,
Se analizarán algoritmos de reconstrucción de poses y localización de objetos 3D basados en imágenes de última generación para imágenes monoculares.
configuraciones de visión estéreo y entrada RGB-D. Cada modalidad de entrada se revisará inicialmente independientemente del dominio antes de que se destaquen los enfoques específicos del alcance quirúrgico. En la segunda parte, se describirán los enfoques actuales de la navegación quirúrgica basada en AR.

2.1. Reconstrucción de poses basada en imágenes

Tradicionalmente, la reconstrucción de poses a partir de imágenes monoculares es
logrado al encontrar un conjunto de correspondencias de puntos 2D/3D entre el espacio de la imagen y la geometría del mundo real. Una vez que se establecen las correspondencias, el algoritmo Perspective-n-Point (PnP) (Lepetit et al., 2009) se puede utilizar en combinación con Random Sample Consensus (RANSAC) (Fischler y Bolles, 1981)
para determinar la pose de un objeto. Los enfoques tradicionales obtuvieron las correspondencias de puntos utilizando características de la imagen como SIFT (Lowe, 2004) o SURF (Bay et al., 2006). Estas características, sin embargo, requieren objetos texturizados para detectar de manera robusta el mismo punto en múltiples imágenes bajo condiciones de iluminación o vista cambiantes. anglos. En los enfoques más recientes, las correspondencias de puntos 2D/3D se encuentran utilizando técnicas de aprendizaje profundo, como las redes neuronales de codificador-descodificador (Pavlakos et al., 2017; Peng et al., 2019;
Hu et al., 2018). Debido al mapeo no lineal de imágenes 2D a puntos del mundo 3D y debido a ocluidos o truncados

objetos en la imagen, los puntos clave rara vez se retroceden directamente. PVNet (Peng et al., 2019), por ejemplo, maneja la oclusión y el truncamiento mediante la regresión de un vector de dirección 2D para cada píxel que pertenece a un objeto que sirve como entrada para la votación de puntos clave. Otros enfoques (Tekin et al., 2018; Hu et al., 2018; Rad y Lepetit, 2017) intentan encontrar las ocho proyecciones de imágenes de las esquinas del cuadro delimitador 3D
en lugar de puntos en el objeto mismo. Sin embargo, debido a la separación de la detección de puntos clave y la reconstrucción de poses usando el algoritmo PnP, estos enfoques pueden carecer de robustez ya que no se pueden entrenar de forma integral. Se han propuesto modelos de reconstrucción de pose entrenables de extremo a extremo (Xiang et al., 2017; Hu et al., 2020) como alternativa, pero son demasiado lentos para su uso en tiempo real o no son aplicables de forma genérica. Los datos de la imagen monocular son también se usa ampliamente en el campo quirúrgico, ya que se puede obtener fácilmente de sistemas de cámaras quirúrgicas como los endoscopios. Las redes de codificador-decodificador también han demostrado su eficacia con estos datos, ya sea para segmentar herramientas quirúrgicas (Ni et al., 2019; Shvets et al., 2019) o para encontrar las posiciones conjuntas de herramientas quirúrgicas articuladas (Kurmann et al., 2017; Du et al., 2018). Otros enfoques también incorporaron la mano del cirujano y, por lo tanto, probaron el seguimiento de herramientas manuales sin marcadores con varias redes neuronales de última generación (Hein et al., 2021).

Inspiradas en la visión binocular de los humanos para percibir información de profundidad, las configuraciones de cámaras estéreo aprovechan las relaciones geométricas conocidas entre dos cámaras para determinar la posición 3D de los objetos en una escena. Para reconstruir la ubicación 3D de un objeto, se deben obtener correspondencias de imágenes 2D/2D que representen los mismos puntos del mundo 3D en cada par de imágenes estéreo. Residencia en estas correspondencias y los parámetros de la cámara, la posición 3D de un objeto se puede determinar por triangulación. Tradicionalmente, las correspondencias se encontraban extrayendo características de imagen escasa o densa de ambas imágenes y combinándolas en función de un medida de similitud seleccionada, a menudo acelerada y hecha más robusta teniendo en cuenta la geometría epipolar (Zhang et al., 1995; Baumberg, 20 0 0; Deriche et al., 1994; Pritchett y Zisserman, 1998; Scharstein et al., 2001). Una vez que las correspondencias se obtuvieron, la ubicación 3D de un punto en particular se puede reconstruir utilizando un método de triangulación como el punto medio directo, Direct Linear Transform (DLT), o explotando el es- matriz esencial. En los enfoques basados en datos, las redes de aprendizaje profundo han demostrado ser más efectivas porque pueden crear mapas de características consistentes procesando simultáneamente un par de imágenes estéreo.
a través de dos ramas de entrada idénticas de capas convolucionales. Después de la concatenación de los mapas de características izquierdo y derecho, la información 3D se puede reconstruir usando un decodificador dedicado (Xie et al., 2019) o una red de propuesta de región (RPN) (Li et al., 2019). Él
Este último también ha propuesto una forma elegante de garantizar la correspondencia en pares de imágenes estéreo mediante la regresión de un cuadro delimitador de unión.
encapsulando el cuadro delimitador izquierdo y derecho después de la superposición de ambas imágenes estéreo. El uso más destacado del estéreo. redes neuronales en el dominio quirúrgico es la endoscopia y la cirugía microscópica, donde se han establecido dispositivos con óptica estereoscópica. Probst et al. (2018) utilizaron dos redes de codificador-decodificador consecutivas para detectar puntos clave de una punta robótica en ambas imágenes de microscopio estereoscópico, individualmente. Otros enfoques evaluaron diferentes algoritmos de reconstrucción 3D en imágenes estereoscópicas endoscópicas simuladas y reales (Parchami et al., 2014).

Una tercera modalidad de entrada que es ampliamente utilizada en el campo de AR

son datos RGB-D que proporcionan información de profundidad junto con una imagen RGB regular (Zhang and Cao, 2017; Kehl et al., 2016; Tan et al., 2017; Sridhar et al., 2016; Whelan et al., 2013). De manera similar a las imágenes monoculares, la mayoría de los enfoques de reconstrucción de poses de objetos se basan en
Los datos RGB-D se basan en características o descriptores que se utilizan para inferir la pose de un objeto. Mientras que algunos utilizan funciones convencionales como
como SURF (Wang and Guo, 2017) otros enfoques proponen un conjunto de características distintas en los datos RGB-D usando aprendizaje profundo (Zeng et al., 2017; Bo et al., 2014; Kehl et al., 2016). la conversión de características a posar se resuelve de manera muy diferente, que van desde 6D object plantean votar a la minimización de una función de energía. Otros enfoques utilizan bosques aleatorios o máquinas de vectores de soporte para hacer retroceder la pose del objeto a partir de una representación intermedia que se origina en una red neuronal (Brachmann et al., 2014; Zia et al., 2017; Schwarz et al., 2015). Además de los métodos diseñados específicamente para datos RGB-D, también existen extensiones para la pose de objetos monoculares. algoritmos de reconstrucción que utilizan información de profundidad para el refinamiento de la pose (Xiang et al., 2017). Los datos RGB-D, sin embargo, tienen inconvenientes significativos, como una mala relación señal-ruido, datos incompletos, un rango de medición limitado, demandas altas de memoria y baja resolución, especialmente para objetos pequeños. En consecuencia, las aplicaciones en el campo médico se limitan predominantemente al análisis de la postura de personas sentadas (Liu et al., 2017) o acostadas (Wu et al., 2020), así como a la detección de personal médico en el quirófano (Kadkhodamohammadi , 2016).

Las consideraciones anteriores alentaron el uso de datos de imagen estéreo

en este trabajo. Los enfoques estéreo explotan toda la información geométrica disponible, brindan información de profundidad más precisa en comparación con los enfoques monoculares (Li et al., 2019), requieren poca memoria y son adecuados para detectar objetos pequeños o puntos clave.

2.2. Navegación quirúrgica basada en AR

AR se ha descrito como una tecnología con el potencial de cambiar radicalmente la cirugía al proporcionar información directamente en el campo de visión del cirujano (Jud et al., 2020). Una ventaja sobre
sistemas de visualización convencionales es la posibilidad de mostrar modelos 3D muy complejos espacialmente precisos en un entorno del mundo real. Los dispositivos AR montados en la cabeza logran la autolocalización requerida sin necesidad de un sistema de seguimiento externo mediante el uso de métodos de localización y seguimiento simultáneos (SLAM) (Durrant-Whyte y Bailey, 2006). En cirugía ortopédica, la visualización mejorada con AR se ha investigado principalmente para anatomía rígida como el hueso (Burström et al., 2019; Jud et al., 2020; Laverdière et al., 2019; Elmi-Terander et al., 2019).), pero también para mostrar estructuras de tejido blando ocultas como una superposición en la anatomía de oclusión real.

Salah et al. (2011) presentó un estudio in vitro en el que se reconstruyeron imágenes de modelos 3D de la columna vertebral, incluidas vértebras, disco, los nervios y la médula espinal se superpusieron in situ a un torso sintético. Se han descrito estudios de prueba de concepto similares para cirugías de columna mínimamente invasivas (Nguyen et al., 2019; Deib et al., 2018), así como para cadera (Chen et al., 2015) y tumores ortopédicos (Cho et al. ., 2018) cirugía. Recientemente, AR se ha propuesto en el contexto de aplicaciones con mayores demandas de precisión, como la navegación quirúrgica. Por ejemplo, propusimos un enfoque de navegación quirúrgica en el que se usó el dispositivo AR HoloLens para navegar por la colocación de tornillos pediculares (Liebmann et al., 2019) y cincelado en cirugías de cadera (Hoch et al., 2021; Ackermann et al., 2021) . Otras soluciones utilizan fluoroscopia con arco en C (Andress et al., 2018)
o marcadores ópticos (Abe et al., 2013; Liu et al., 2018) para registrar un plan preoperatorio a la anatomía intraoperatoria que permite navegar por puntos de entrada o trayectorias de herramientas quirúrgicas en vertebroplastia percutánea y resuperficialización de cadera, respectivamente. ARKANSAS-
Los enfoques basados en navegación quirúrgica brindan a los cirujanos información adicional en situaciones donde la percepción humana es limitada y, en consecuencia, traducen procedimientos altamente complejos en tareas más simples, como se demuestra en este enfoque centrado en el usuario.
(Brendle et al., 2020). Esto no solo reduce el riesgo de complicaciones intraoperatorias graves, sino que también aumenta notablemente la precisión del resultado quirúrgico. Los estudios informaron sobre una precisión general de 0,8 mm–2,8 mm y 1,0° - 3,4° (ver Chen et al. (2015);
Burström et al. (2019); Liebmann et al. (2019) para más detalles) en la ejecución de tareas de perforación quirúrgica y reducción ósea.

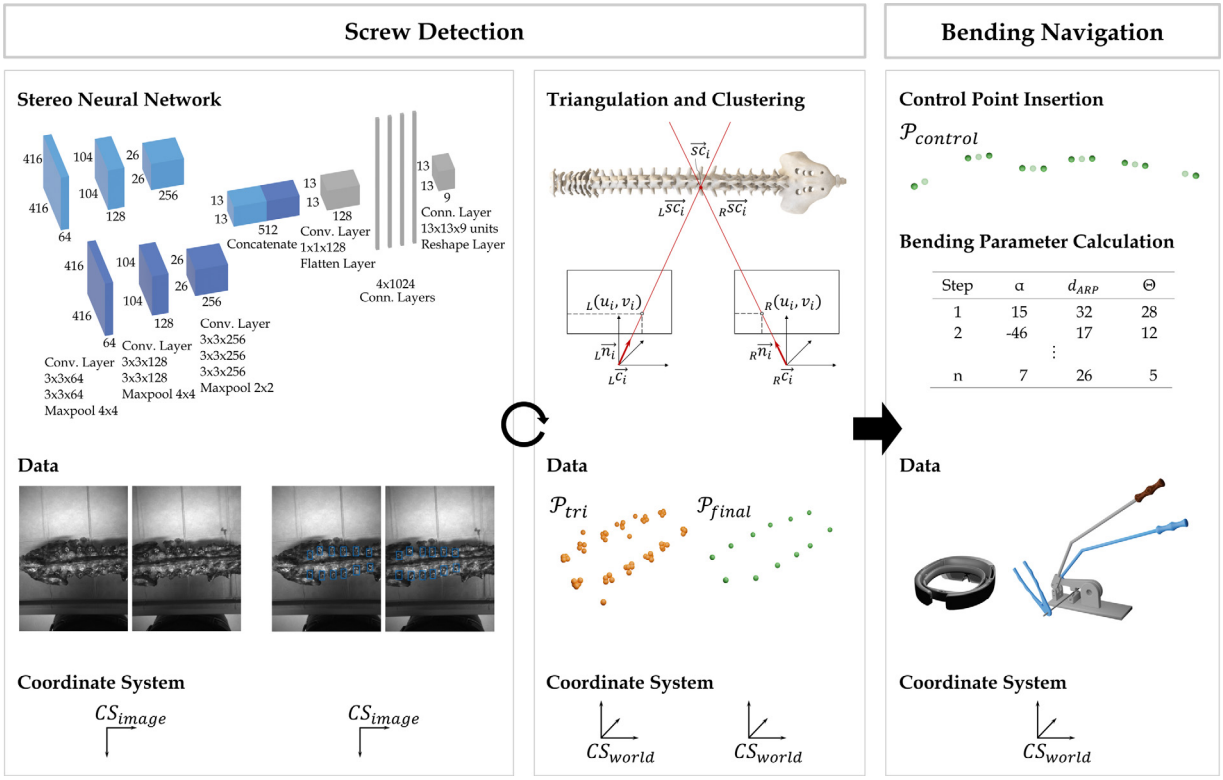


Fig. 2. La canalización de navegación de flexión de varilla basada en AR propuesta que consiste en la detección de la cabeza del tornillo a partir del flujo de imágenes estéreo, la reconstrucción 3D de las posiciones de la cabeza del tornillo, la generación de instrucciones de flexión y la navegación basada en AR de la flexión del implante de varilla. Detección de tornillo (columna izquierda): arquitectura de red (arriba); datos de entrada y salida (centro), descritos en los sistemas de coordenadas de imagen correspondientes. Triangulación (columna central): Método de triangulación (arriba); candidatos de cabeza de tornillo triangulada P tri y posiciones de tornillo agrupadas P final (centro), dadas en coordenadas mundiales. Tenga en cuenta que la agrupación en clústeres es un proceso iterativo que tiene en cuenta los resultados de la triangulación de varios fotogramas de imágenes estéreo. Navegación de plegado (columna derecha): Los puntos de control P controlan la inserción y el cálculo de instrucciones de plegado (arriba); Navegación de flexión basada en AR (centro), descrita en coordenadas mundiales.

Sin embargo, los enfoques AR intraoperatorios actuales visualizan un objetivo específico o una desviación de ese objetivo sin guiar al usuario.
sobre cómo se puede lograr la situación deseada. Por lo tanto, los métodos aún dependen en gran medida de la destreza del usuario. Otros dominios de aplicación ya han aprovechado la RA en este contexto al brindar orientación paso a paso en el mantenimiento, el ensamblaje o la capacitación (De Amicis et al., 2018; Webel et al., 2013; Westerfield et al. ., 2015; Zhu et al., 2014; Wang et al., 2020; Sorko y Brunnhofer, 2019). De Amicis et al. (2018) creó un manual interactivo para
ensambles complejos proporcionando un paso a paso basado en AR
secuencia de orientación. Al rastrear los dedos del usuario con un rastreador de movimiento Leap, incluso sugieren una trayectoria de movimiento de la mano en lugar de
que simplemente proponer la ubicación deseada de una pieza en el ensamble. Sorko y Brunnhofer (2019) analizaron el potencial de la tecnología AR para la educación y la formación en escenarios industriales. Descubrieron que no solo se acortaron los tiempos de proceso, sino que también se redujeron las tasas de error. El objetivo de nuestro estudio fue traducir este concepto guía secuencial paso a paso a la cirugía haciendo
Tareas dependientes de la destreza más simples y precisas a través de la automatización y la guía AR.

3. Métodos

Nuestra tubería propuesta para la navegación AR sin marcadores del proceso de doblado de varillas se muestra en la Fig. 2 y consta de dos principales
componentes, a saber, la detección de tornillo (Sección 3.1) y la subsiguiente navegación de flexión (Sección 3.2). Las entradas a la canalización son imágenes estéreo en escala de grises que se transmiten continuamente desde
las dos cámaras ambientales frontales de HoloLens de primera generación con una resolución de 480 × 640 píxeles. Las imágenes estéreo se alimentan a las dos ramas de la red neuronal estéreo.
(Sección 3.1.1) que determina un par correspondiente de límites
recuadros en ambas imágenes estereoscópicas para cada detección de cabeza de tornillo pedicular.
Los centros de los cuadros delimitadores correspondientes se asignan a espacio 3D por triangulación y procesado por un enfoque basado en agrupamiento que se utiliza para recopilar y refinar las posiciones 3D a lo largo del tiempo a medida que se procesan más fotogramas estéreo (Sección 3.1.2). Las estimaciones puntuales resultantes se convierten luego en instrucciones de doblado de varillas.
(Sección 3.2.1) y proporcionado al cirujano en forma de un banco de flexión aumentado (Sección 3.2.2). Los detalles sobre la evaluación experimental se pueden encontrar en la Sección 3.3.

3.1. Detección de tornillo

Nuestra red neuronal estéreo propuesta amplía la noción bien establecida de un cuadro delimitador de unión (Li et al., 2019) para asociar detecciones estéreo mediante la representación de salida unificada de
YOLO (Redmon y Farhadi, 2018). El modelo resultante se puede entrenar de extremo a extremo y se ejecuta en tiempo real.

3.1.1. Arquitectura de red y entrenamiento

Ambas ramas de entrada de la red están configuradas de forma idéntica. Cada rama consta de tres bloques convolucionales compuestos por
una serie de capas convolucionales con filtros de 3 × 3 que se completan con una agrupación máxima y una capa de abandono (ver Fig. 2). Las activaciones de cada capa convolucional son posprocesadas por un lote
capa de normalización. El número de filtros se duplica para cada bloque convolucional y los pesos de las capas convolucionales son
compartida por ambas ramas de la red. Esta estrategia permite la generación de mapas de características consistentes para la entrada izquierda y derecha. imagen, respectivamente. En la siguiente parte de la red, los dos mapas de características se concatenan antes de que una capa convolucional reduzca

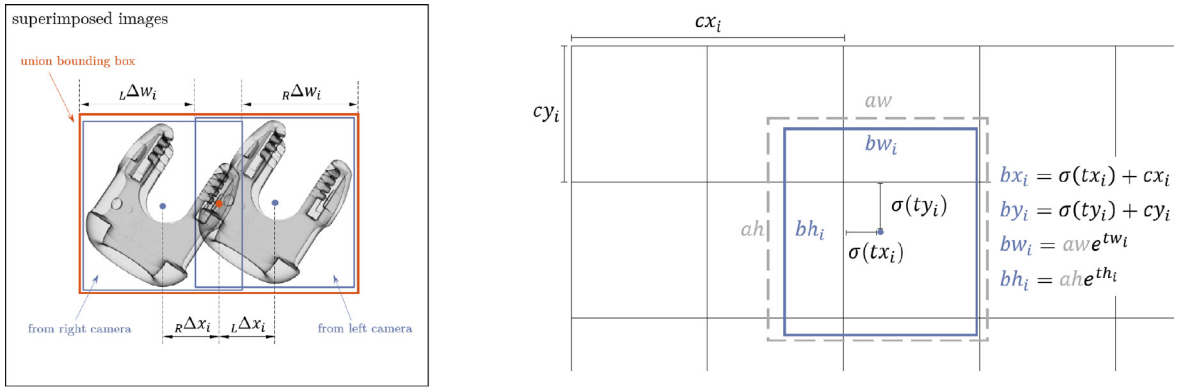


Fig. 3. (a) Los cuadros delimitadores de la cabeza de un tornillo visibles en la imagen izquierda y derecha se muestran en azul. Después de la superposición de la imagen izquierda y derecha, se asigna un cuadro delimitador de unión (rojo) que contiene los cuadros delimitadores izquierdo y derecho de una cabeza de tornillo. (b) Esta figura está adaptada de Redmon y Farhadi (2018). Los valores t xi , t yi , t wi y thi codifican el cuadro delimitador. Todos los valores se definen en relación con la esquina superior izquierda de la celda respectiva (cxi, cyi). La función sigmoidea σ normaliza las activaciones en el rango [0,1]. El ancho y la altura del ancla (aw y ah) sirven como antes para el cuadro delimitador. (Para la interpretación de las referencias al color en la leyenda de esta figura, se remite al lector a la versión web de este artículo).

la dimensionalidad del tensor resultante usando filtros 1 × 1. Luego, cuatro capas totalmente conectadas con 1024 unidades cada una hacen retroceder el tensor de salida final que se lleva a la forma deseada mediante una capa de remodelación. La función de activación para todas las capas convolucionales y densas es una ReLU con fugas con una pendiente de 0,1 excepto para la última capa densa que usa mapeo lineal.

El tensor de salida de la red tiene la forma 13 × 13 × 9
y contiene la información codificada para reconstruir todos los cuadros delimitadores en ambas imágenes estéreo. Como se discutió en la Sección 2, encontrar correspondencias mediante la asociación de objetos detectados en las imágenes izquierda y derecha es una tarea desafiante. Nuestro enfoque aborda este
problema superponiendo las imágenes de entrada izquierda y derecha para crear un cuadro delimitador de unión para cada tornillo que contiene tanto el cuadro delimitador de la imagen izquierda como la derecha, respectivamente. Este concepto se ilustra para una sola cabeza de tornillo pedicular en la Fig. 3a. El tensor de salida de la red se puede interpretar como una cuadrícula de 13 × 13 que divide la imagen en 169 celdas, cada una de las cuales consta de nueve valores de regresión. Un tamaño de cuadrícula de salida de 13 × 13 fue heurísticamente
determinado como una buena compensación que permite la detección de múltiples objetos vecinos más pequeños, como tornillos, manteniendo la dimensionalidad baja. Cada celda es responsable de detectar cuadros delimitadores de unión cuyo centro se encuentra dentro de los límites de la célula. Cada detección γi (i = 1, . . . , N det , donde N det denota el número de detecciones de cabeza de tornillo) y, en consecuencia, cada cuadro delimitador de unión que se encuentra en cada par de imágenes estéreo en la cuadrícula de 13 × 13 es en - codificado por un total de nueve parámetros retrocedidos organizados en tres grupos:

$$\gamma_i = \left[\underbrace{ts_i}_{\text{presencia}}, \underbrace{tx_i, ty_i, tw_i, th_i}_{\text{caja de union}}, \underbrace{{}_L t\Delta x_i, {}_L t\Delta w_i, {}_R t\Delta x_i, {}_R t\Delta w_i}_{\text{correccion stereo}} \right]$$

El primer parámetro tsi indica si un tornillo y, en consecuencia,
el centro de un cuadro delimitador de unión se encuentra en la celda de cuadrícula respectiva. Este parámetro es una variable binaria para entrenamiento pero
necesita exceder un valor determinado experimentalmente de 0.5 para sugerir la presencia de un tornillo durante la inferencia. Las siguientes cuatro entradas t xi , t yi , t wi , thi definen la ubicación precisa de un cuadro delimitador de unión, así como su ancho y alto. Supongamos que cada celda de la
cuadrícula de 13 × 13 tiene unidad de ancho y alto y que la parte superior izquierda
La distribución de una celda de cuadrícula se puede describir mediante los dos valores (cxi, cyi), como se muestra en la Fig. 3 b. En lugar de retroceder el límite de la unión
ubicación del cuadro en coordenadas de píxeles globales, todos los refinamientos se describen en relación con la ubicación de la celda (cxi, cyi) dentro del 13 × 13
rejilla, donde ocurrió la detección. El cuadro delimitador de unión

Los parámetros bxi , byi , bwi , bhi se obtienen entonces por

$$bx_i = \sigma(tx_i) + cx_i$$

byi = σ (tyi) + cyi
bwi = aw · e twi
bhi = ah · e thi
donde σ denota la función sigmoidea. Los parámetros aw y ah son valores ancla que introducen conocimientos previos sobre los cuadros delimitadores de la unión. Esta información previa se obtiene promediando los cuadros delimitadores de la unión de verdad de tierra etiquetados manualmente para proporcionar una buena estimación inicial que se corrige con los términos exponenciales (consulte Redmon y Farhadi (2018) para obtener una descripción detallada del concepto de cuadro de anclaje). Tenga en cuenta que los anclajes tienen un tamaño predefinido y no dependen de la detección actual. Se requiere la función sigmoidea σ para mapear los parámetros de regresión txi y tyi en el rango [0,1] para garantizar que el centro del cuadro delimitador permanecerá en la celda de cuadrícula predicha.
El tercer grupo de parámetros se refiere a la corrección estéreo que determina las compensaciones desde el cuadro delimitador de unión a los respectivos cuadros delimitadores en la imagen izquierda y derecha. Asumiendo cámaras rectificadas, solo es necesario retroceder el desplazamiento horizontal y la corrección de ancho para la imagen izquierda y derecha (ver Fig. 3 a).
Esto da como resultado los cuatro parámetros finales del descriptor de detección L t ??xi , L t ??wi , R t ??xi , R t ??wi . Nótese que un prescripto • indica a partir de ahora que se puede aplicar un término a la imagen izquierda y derecha, respectivamente. El descriptor de detección se convertirá en compensaciones absolutas de la siguiente manera:

$$\bullet\Delta x_i = \bullet a\Delta x \cdot e^{\bullet t\Delta x_i}$$
$$\bullet\Delta w_i = \bullet a\Delta w \cdot e^{\bullet t\Delta w_i}$$

Tenga en cuenta que los parámetros •a fix y •a fiw son valores ancla que se encontraron al promediar las compensaciones horizontales observadas y las correcciones de ancho en los datos reales del terreno, similares a los valores ancla aw y ah anteriores. Con esta representación, cada celda de la cuadrícula puede detectar exactamente una cabeza de tornillo. Los cuadros delimitadores • (xi , yi , wi , hi) en los pares de imágenes estéreo eventualmente se encuentran de la siguiente manera:

$$\begin{aligned} {}_L x_i &= bx_i + {}_L\Delta x_i & {}_R x_i &= bx_i - {}_R\Delta x_i \\ {}_L y_i &= by_i & {}_R y_i &= by_i \\ {}_L w_i &= bw_i - {}_L\Delta w_i & {}_R w_i &= bw_i - {}_R\Delta w_i \\ {}_L h_i &= bh_i & {}_R h_i &= bh_i \end{aligned}$$

Las detecciones de puntos finales en el espacio de píxeles • (ui, vi) se encuentran transformando el centro de los cuadros delimitadores de espacio de cuadrícula a

espacio de píxeles. Esta transformación consiste en dividir • (xi , yi) por 13, resultando en coordenadas normalizadas, y sucesivas multiplicaciones por el ancho y alto de la imagen original, respectivamente.

Nuestra red se implementó en TensorFlow y se entrenó con un conjunto de datos obtenido de los experimentos ex vivo descritos en la Sección 3.3.1. Para homogeneizar el conjunto de datos, todas las imágenes se redimensionaron a una resolución de 416 × 416 píxeles y se normalizaron. Después de la inicialización del peso aleatorio, la red neuronal estéreo se entrenó desde cero durante 10 0 0 épocas con un tamaño de lote de 16. La tasa de aprendizaje se estableció inicialmente en 10 −3 y se redujo a 10 −4 después de 750 épocas y a 10 −5 durante las últimas 100 épocas. El entrenamiento tomó aproximadamente 14
horas en un NVIDIA Quadro RTX 60 0 0. Para generalizar mejor a los datos no vistos, las imágenes estéreo se aumentaron sobre la marcha para el entrenamiento. Con este fin, se examinaron 15 estrategias de aumento diferentes con diversas combinaciones de técnicas de aumento como como cambios de brillo y contraste, desenfoque, ecualización de histograma, escalado, volteo vertical de la imagen y traslación vertical. Después de una cuidadosa evaluación, una combinación de traslación vertical con una probabilidad del 50 % con escalamiento posterior o adaptación de contraste produjo el mejor rendimiento de detección.

3.1.2. Triangulación y agrupamiento

Dadas las detecciones i correspondientes • (ui, vi) en un par de imágenes estéreo, la posición 3D del i-ésimo tornillo candidato −sc → puedo ser determinada utilizando el método del punto medio del vector de la siguiente manera (más información
y la nomenclatura se puede encontrar en la Fig. 2). Sean L n→ i y R n→ i los vectores directores normalizados de los rayos de los respectivos centro de la cámara L → i y R c→ i a las detecciones L (ui, vi) y R (ui, vi). Como los rayos normalmente no se cruzan, estamos interesados en determinando los puntos $\vec{sc_i}$ y a la izquierda y el rayo derecho cuales son más cerca uno del otro:

$$ {}_L\vec{sc_i} = {}_L\lambda_i \cdot {}_L\vec{n}_i $$
$$ {}_R\vec{sc_i} = \left({}_R\vec{c}_i - {}_L\vec{c}_i\right) + {}_R\lambda_i \cdot {}_R\vec{n}_i $$

La siguiente ecuación se puede enunciar teniendo en cuenta que

$_L$ sc → i −R sc → i tiene que ser perpendicular a ambos rayos para garantizar la menor distancia. Proyectando ambos rayos uno sobre el otro y dado que L $\vec{sc_i}$ y $_R$ sc→i coinciden en esta situación proyectada da como resultado:

$$ {}_L\lambda_i \cdot \left({}_L\vec{n}_i \cdot {}_L\vec{n}_i\right) = \left({}_R\vec{c}_i - {}_L\vec{c}_i\right) \cdot {}_L\vec{n}_i + {}_R\lambda_i \cdot \left({}_R\vec{n}_i \cdot {}_L\vec{n}_i\right) $$
$$ {}_L\lambda_i \cdot \left({}_L\vec{n}_i \cdot {}_R\vec{n}_i\right) = \left({}_R\vec{c}_i - {}_L\vec{c}_i\right) \cdot {}_R\vec{n}_i + {}_R\lambda_i \cdot \left({}_R\vec{n}_i \cdot {}_R\vec{n}_i\right) $$

Solving for $_L\lambda_i$ and $_R\lambda_i$ results in:

$$ {}_L\lambda_i = \frac{\left({}_R\vec{c}_i - {}_L\vec{c}_i\right) \cdot {}_L\vec{n}_i - \left({}_R\vec{c}_i - {}_L\vec{c}_i\right) \cdot {}_R\vec{n}_i \cdot \left({}_L\vec{n}_i \cdot {}_R\vec{n}_i\right)}{1 - \left({}_L\vec{n}_i \cdot {}_R\vec{n}_i\right)^2} $$
$$ {}_R\lambda_i = \frac{\left({}_R\vec{c}_i - {}_L\vec{c}_i\right) \cdot {}_L\vec{n}_i \cdot \left({}_L\vec{n}_i \cdot {}_R\vec{n}_i\right) - \left({}_R\vec{c}_i - {}_L\vec{c}_i\right) \cdot {}_R\vec{n}_i}{1 - \left({}_L\vec{n}_i \cdot {}_R\vec{n}_i\right)^2} $$

Finalmente $\vec{sc}_i$ se determina por interpolación lineal de $_L\vec{sc_i}$ and $_R\vec{sc_i}$ y colocado en el conjunto puntual de puntos triangulados $\mathcal{P}_{tri}$.

Cada par de imágenes estéreo procesadas proporciona N det posibles candidatos a tornillos que se expresan en un marco de coordenadas mundiales 3D y se almacenan en un conjunto de puntos P tri. El algoritmo SLAM patentado de HoloLens garantiza la coherencia espacial entre los puntos. El objetivo de la rutina de agrupamiento subsiguiente es condensar las ubicaciones candidatas para los tornillos entrantes −sc→ i de P tri sobre la marcha en un conjunto de posiciones de tornillos agrupados P final en función del número de tornillos deseados N tornillos antes de la operación. El algoritmo de agrupamiento funciona de la siguiente manera (ver Algoritmo 1 para más detalles).

Algorithm 1

$d_{thresh} \leftarrow 2.0$
$i_{det} \leftarrow 0$
$i_{clu} \leftarrow 0$
$\mathcal{P}_{cand} \leftarrow \emptyset$
$running \leftarrow True$
while *running* **do**
  $p_{curr} \leftarrow$ get $\vec{sc}_{i_{det}}$ from $\mathcal{P}_{tri}$
  $i_{det} \leftarrow i_{det} + 1$
  if $\mathcal{P}_{cand}$ is $\emptyset$ **then**
    $\mathcal{P}_{i_{clu}} \leftarrow \emptyset \cup p_{curr}$
    $\mathcal{P}_{cand} \leftarrow \mathcal{P}_{cand} \cup \mathcal{P}_{i_{clu}}$
    $i_{clu} \leftarrow i_{clu} + 1$
  else
    $d_{\min} \leftarrow \infty$
    $\mathcal{P}_{closest} \leftarrow \emptyset$
    for each $\mathcal{P} \in \mathcal{P}_{cand}$ **do**
      $d = \text{closest\_dist}(p_{curr}, \mathcal{P})$
      if $d < d_{\min}$ **then**
        $d_{\min} \leftarrow d$
        $\mathcal{P}_{closest} \leftarrow \mathcal{P}$
    if $d_{\min} < d_{thresh}$ **then**
      $\mathcal{P}_{closest} \leftarrow \mathcal{P}_{closest} \cup p_{curr}$
    else
      $\mathcal{P}_{i_{clu}} \leftarrow \emptyset \cup p_{curr}$
      $\mathcal{P}_{cand} \leftarrow \mathcal{P}_{cand} \cup \mathcal{P}_{i_{clu}}$
      $i_{clu} \leftarrow i_{clu} + 1$
  $c_{confirmed} \leftarrow 0$
  for each $\mathcal{P} \in \mathcal{P}_{cand}$ **do**
    if $|\mathcal{P}| \geq 100$ points **then**
      $c_{confirmed} \leftarrow c_{confirmed} + 1$
  $running \leftarrow c_{confirmed} < N_{screws}$
$\mathcal{P}_{final} \leftarrow \emptyset$
for each $\mathcal{P} \in \mathcal{P}_{cand}$ **do**
  $p_{mean} \leftarrow$ center($\mathcal{P}$)
  $\mathcal{P}_{final} \leftarrow \mathcal{P}_{final} \cup p_{mean}$

La primera candidata a tornillo entrante p_{curr} de $\mathcal{P}_{tri}$ se almacena en un nuevo conjunto de puntos $\mathcal{P}_{i_{clu}}$ que se agrega al conjunto de candidatos de racimo $\mathcal{P}_{cand}$. Cada punto entrante p actual después de eso se agrega al grupo existente más cercano P más cercano ∈ P cand, si la distancia al centro del grupo más cercano es menor que una distancia d umbral determinada empíricamente de 2,0 cm. De lo contrario, el punto p curr es la semilla de que se suma a P c y. El procedimiento termina tan pronto como se encuentran N grupos de tornillos que están soportados por 100 puntos individuales. Los centros de conglomerados finales se determinan encontrando el centro de cada punto establecido en Pcand y almacenándolo en Pfinal. Estas estimaciones finales se presentan al cirujano para su confirmación visual. En caso de detecciones incorrectas, se reinicia la detección de tornillos.

Este algoritmo no solo reduce un juego de tornillos potencialmente ruidoso señala a los candidatos en un número distinto de estimaciones, pero también elimina de manera eficiente los valores atípicos debido a la falta de apoyo de otros candidatos puntos. Finalmente, se aplica el Análisis de Componentes Principales para separar todos los puntos en Pf inal en un conjunto de puntos anatómicamente izquierdos.

$_L\mathcal{P}_{final}$ y puntos correctos $_R\mathcal{P}_{final}$, respectivamente y ordenar todos los puntos de cranial ($j = 1$) a caudal ($j = N_{screws}/2$).

3.1.3. Transmisión

Para permitir tarifas interactivas, algunos de los cálculos deben realizarse en una estación de trabajo de alta gama en lugar de a bordo de HoloLens. Con este fin, los datos de imagen estéreo se transmiten desde HoloLens a una PC mediante una aplicación C++ personalizada. en la estación de trabajo-

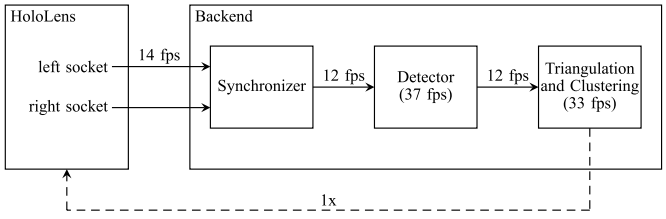


Fig. 4. Data workflow: The HoloLens is continuously streaming image data to the backend computer that infers bounding boxes on the stereo images and subsequently triangulates the 3D detections. The clustering routine runs continuously based on incoming 3D detections until the required number of screws are found. Upon convergence, the clustered final screw estimates are sent back once (dashed line) to the HoloLens for display and verification.

tion, a Python backend was implemented to calculate the 3D positions of the pedicle screws which are eventually sent back to the HoloLens for display and verification. The data flow is visualized in Fig. 4.

The following components were implemented:

HoloLens (C++) This application establishes a connection for each targeted camera to a separate workstation. We chose TCP as the connection protocol to achieve high reliability, congestion control, and lossless data transmission. In our setup, handshaking takes approximately 90 ms. After establishing the connections, each socket asynchronously sends the latest image and camera pose information at image capture. Duplicate deletion ensures maximum information flow. For accurate 3D reconstruction, camera distortion parameters are sent once per connection.

Synchronizer The synchronizer receives images at a rate of 14 fps and synchronizes the incoming images from the right and left socket based on their timestamp at capture. A time difference of up to 20 ms was accepted for synchronization which resulted in an output frame rate of 12 fps.

Detector The detector receives synchronized stereo images and infers all corresponding bounding boxes using the stereo neural network described in Section 3.1.1. This could be achieved at a frame rate of 37 fps, but is restricted by the incoming frame rate of 12 fps.

Triangulation and Clustering The detections are then unprojected and triangulated, and subsequently passed to the clustering routine. Once the algorithm terminates, the 3D positions are sent back once to the HoloLens for point display and surgeon verification. Unprojection and triangulation is also performed at 12 fps, even though 33 fps would be the measured maximum assuming 37 fps from the detection step.

3.2. Bending navigation

Once the 3D positions of the pedicle screws $p_j \in {}_{\bullet}\mathcal{P}_{final}$ have been obtained, the bending instruction is generated to eventually guide surgeons with the HoloLens. Instead of simply providing a visualization of the desired rod shape, we propose a more intuitive process where every bending step is navigated individually. The following section is structured into two parts. Firstly, the bending parameters are introduced and their calculation is explained in Section 3.2.1. Secondly, the translation of bending parameters to the intraoperative navigation is described in Section 3.2.2.

3.2.1. Bending parameter

Each bending step is characterized by a set of bending parameters, as depicted in Fig. 1a. To adjust the rod position and orientation, the implant needs to be shifted along its main axis by *dARP* and rotated by α. The bending angle of the rod implant β is proportional to the angular displacement of the bending bench lever Θ.

The pedicle screw head tulip, where the rod will eventually be mounted, has an opening that is only 0.1 mm wider than the diameter of the rod to guarantee a strong rigid postoperative connection. This implies that the rod has to be straight in the positions where it will be mounted into the pedicle screw heads. To ensure straight rod segments between the screw heads, each screw head p_j in ${}_{\bullet}\mathcal{P}_{final}$ is replaced by two equidistant control points and added to the respective set of control points ${}_{\bullet}\mathcal{P}_{control}$.

$$ {}_{\bullet}\mathcal{P}_{control} \leftarrow \vec{p}_j \pm \mu \cdot \frac{\left(\vec{p}_{j+1} - \vec{p}_{j-1}\right)}{\left\|\left(\vec{p}_{j+1} - \vec{p}_{j-1}\right)\right\|} \forall p_j; j = 2, \ldots, |{}_{\bullet}\mathcal{P}_{final}| - 1 $$

where μ is a heuristically determined parameter that was set to 7.5 mm. From this set of control points, all bending parameters can be calculated for each bending step. The k^{th} bend is characterized by the required bending angle of the rod β_k, the axial reorientation angle α_k and the distance by which the rod needs to be advanced $dARP_k$ as illustrated in Fig. 1a. Each **bending angle** β_k is found by iterating over ${}_{\bullet}\mathcal{P}_{control}$ to determine the angles using the dot product:

$$ \beta_k = \arccos\left(\frac{p_{k+1} - p_k}{\|p_{k+1} - p_k\|} \cdot \frac{p_{k-1} - p_k}{\|p_{k-1} - p_k\|} \right) \forall p_k; k = 2, \ldots, |{}_{\bullet}\mathcal{P}_{control}| - 1 $$

The **axial reorientation angle** α_k for the k^{th} bend is calculated by taking four control points into account and initially generating the following three vectors:

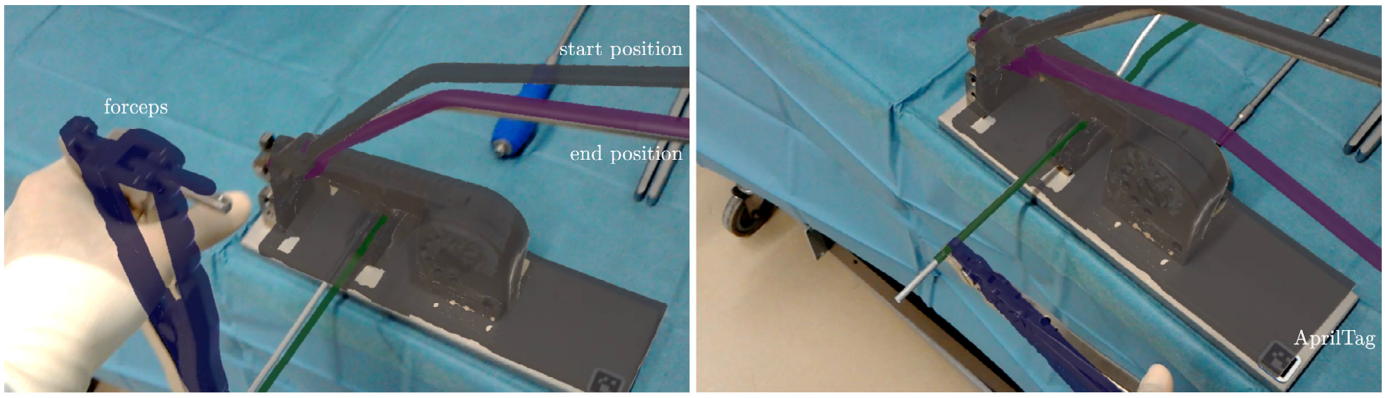
$$ \left.\begin{aligned} {}_{L}\vec{n}_k &= \frac{p_{k-1} - p_k}{\|p_{k-1} - p_k\|} \\ {}_{C}\vec{n}_k &= \frac{p_{k+1} - p_k}{\|p_{k+1} - p_k\|} \\ {}_{R}\vec{n}_k &= \frac{p_{k+2} - p_{k+1}}{\|p_{k+2} - p_{k+1}\|} \end{aligned}\right\} \forall p_k; k = 2, \ldots, |{}_{\bullet}\mathcal{P}_{control}| - 2 $$

In a next step, the vectors ${}_{L}\vec{n}_k$ and ${}_{R}\vec{n}_k$ are projected on the plane defined by the normal vector ${}_{C}\vec{n}_k$ resulting in the projected vectors ${}_{L}\tilde{n}_k$ and ${}_{R}\tilde{n}_k$. Lastly, the axial reorientation angle for the k^{th} bending step is found by:

$$ \alpha_k = \arccos({}_{L}\tilde{n}_k \cdot {}_{R}\tilde{n}_k) $$

The **distance** $dARP_k$ that the rod needs to be displaced in the k^{th} bending step is determined by the Euclidean distance between the last and the current control point.

The **lever angle** Θ_k of the bending bench depends on the desired rod bending angle β_k. Any bending can be considered a combination of elastic and plastic deformation of the rod. Small lever angles result in no permanent rod deformation due to elastic deformation. The relationship between any desired rod angle β and any applied lever angle Θ, however, can be approximated to be linear as soon as plastic deformation starts to occur. This results in a transfer function of the form $\beta = f(\Theta) = m \cdot \Theta + t$, where β is the desired bending angle of the rod, Θ corresponds to the difference in lever angle from start to end position of the bend and m and t denote the slope and offset of the linear model, respectively. To derive the model parameters, a rod was randomly bent 10 times at 3cm intervals while the lever displacement *dLever* was recorded using a custom 3D printed fiducial marker and the optical tracking system Fusiontrack 500 (Atracsys SA, Switzerland). Since the end of the lever describes a circular movement with respect to the center of rotation, the relationship between the straight distance traveled by the tip of the lever *dLever* and the resulting difference in lever angle Θ is given by the equation of a chord which



(a) (b)

Fig. 5. Surgeon's perspective. The surgical bending bench is overlayed by a virtual model. (a) The position and orientation of the rod are navigated by the overlay of the forceps as shown in blue. The start position of the lever is indicated in gray and the desired end position is described by Θ_k (see Section 3.2.1) and is marked in purple. (b) Bending step k+1 is navigated by displaying an updated position of the forceps and lever end position. This process is repeated until all bending steps are completed. (For interpretation of the references to color in this Figure legend, the reader is referred to the web version of this article.)

is $dLever = 2 \cdot r \cdot \sin\frac{\Theta}{2}$, where r denotes the straight line distance from lever base to lever tip as depicted in Fig. 1a. The respective resulting bending angles of the rod β were determined from a CT scan to estimate the aforementioned linear transfer function in a least square sense.

3.2.2. AR Guidance

The bending parameters $dARP_k, \alpha_k, \Theta_k$ are translated to the AR-based step-by-step instruction using following workflow:

1. **Bending bench registration** The routine starts with the registration of the bending bench. For registration, the HoloLens detects an AprilTag using its two front-facing environmental cameras, as described by Liebmann et al. (2019), and an overlay of the bending bench is presented to the surgeon. The exact position of the AprilTag relative to the bending bench was derived from a CT scan.
2. **Preparation** In the next step, an overlay of the surgical forceps is displayed. The displayed forceps is a rod gripper instrument (Ref: 03.51.10.0135; Medacta SA, Switzerland) provided as a standard instrument within the spinal fusion instrument set that is specifically designed for holding rods without slippage. The surgeon fixes the real-world forceps to the end of the rod and aligns the position and orientation of the forceps with the AR overlay (see Fig. 5).
3. For the k^{th} bending step
   - 3.1 **Axial reorientation** Providing guidance for the axial orientation of the rod α_k is accomplished by aligning the forceps axially to the presented overlay.
   - 3.2 **Axial displacement** To guarantee the bending of the rod at the correct position, the implant needs to be shifted axially by $dARP_k$. This step is navigated using the same overlay of the forceps as for the axial reorientation. Again, the real-world forceps that are rigidly connected to the rod need to coincide with the presented overlay. An example is shown in Fig. 5, where the target position of the forceps is shown in blue.
   - 3.3 **Lever movement** The navigation of the bending is achieved by showing the start and end positions of the lever of the bending bench, as illustrated in gray and purple in Fig. 5, respectively. While the start position is fixed, the end position of the lever is displayed according to Θ_k.
4. **Inspection** An overlay of the target rod shape is presented to the surgeon upon completion of the bending procedure. The shape of the rod can be verified by visual inspection and adjustments can be made if necessary.

3.3. Ex-vivo experiments

In this section, the experimental setups for collecting training data and for evaluating the performance of our navigation approach are described in Section 3.3.1 and Section 3.3.2, respectively.

3.3.1. Experimental setup for training data acquisition

Training of our neural network was performed using a dataset of annotated stereo images. To this end, a senior spine surgeon instrumented a total of eight cadavers with 10–14 pedicle screws each. Afterwards, six individuals, two of which were orthopedic surgeons, put on the HoloLens to conduct a 30 seconds data acquisition procedure. The participants were asked to move along the cadaver to record many different viewpoints with a changing number of visible screws while trying to keep the distance between their head and the cadaver relatively constant (approximately 40 cm). This distance was heuristically determined to be a good trade-off between screw visibility and precision. However, a larger distance between the surgeon's head and the patient's anatomy would be desirable from a surgical perspective to ensure sterility. We think that larger distances can be reached with the next generation of sensors having a higher image resolution. A custom-made application was used to record the stereo image data of the two front-facing environmental cameras of the HoloLens. The application also provided real-time cadaver-head-distance feedback using the HoloLens' built-in depth sensor. Two of the eight cadavers were reused for additional data collection by randomly placing screws in bone and soft tissue. One cadaver was instrumented with 17 random screws whereas 21 screws were implanted in the other cadaver. The odd number of screws in combination with their random placement should mitigate the risk of our network to overfit to recurring screw placement patterns. A total of 2027 stereo image pairs (4054 single images) were annotated by three individuals by manually drawing a bounding box around each screw visible in both stereo images using an open-source annotation tool[1]. This resulted in 19815 union bounding box annotations (39630 single bounding box annotations).

The evaluation of the model performance was based on a 10-fold cross-validation, where the mean average precision (mAP) for

[1] https://github.com/AlexeyAB/Yolo_mark

intersection-over-union (IoU) thresholds of 0.25 and 0.5 was calculated. The subsequent triangulation into 3D space and the clustering routine (see Section 3.1.2) were evaluated by comparison to CT ground truth (Somatom Edge CT® device (Siemens, Erlangen, Germany), slice thickness 1.0mm, in-plane resolution 0.5×0.5mm). The 3D models of the inserted pedicle screws were extracted using the global thresholding and region growing functionalities of a commercial segmentation software (Mimics Medical, Materialise NV, Belgium). The CAD 3D models of the pedicle screw heads were then aligned to the CT-extracted 3D models using the iterative closest point algorithm (ICP) (Besl and McKay, 1992). The centers of the registered pedicle screw head models served as ground truth. For five different configurations regarding relative position of HoloLens and anatomy, the obtained 3D screw position estimates of individual rods resulting from our proposed method after triangulation and clustering were registered to the ground truth using ICP. The average 3D distance between our estimate and the ground truth rods is reported.

3.3.2. Experimental setup for navigation evaluation

Our approach was evaluated by two senior spine surgeons on four cadavers and compared against two baseline methods in terms of bending time and rebending maneuvers. This resulted in a total of 24 bent rods (4 specimens with 3 techniques each on both sides of the spine). The **bending time** denotes the time from the initial push down of the lever of the bending bench to the moment when the rod could be placed in the pedicle screws. A correction of the rod shape qualified as a **rebending maneuver** if further ex-situ or in-situ bending maneuvers were required after the first implantation attempt.

The first benchmark method is the conventional **freehand** bending approach where both surgeons performed the following procedure:

1. The surgeon inspects the instrumented anatomy and creates a mental bending plan.
2. The surgeon steps away from the anatomy to the bending bench and the bending time is started.
3. After a few bendings, the surgeon moves to the anatomy and visually assesses the precision of the rod.
4. The surgeon moves back to the bending bench and either applies corrections to the bent part of the rod or continues with the bending process.
5. The two previous steps are repeated roughly 4–5 times until the surgeon declares the rod ready for reduction into the anatomy.
6. If the rod does not yet meet clinical requirements, in-situ and ex-situ rebending maneuvers are required and counted.
7. The bending time is stopped after successful insertion of the rod into the pedicle screw heads.

The following procedure was followed by the surgeons for our second baseline method proposed by Wanivenhaus et al. (2019). We refer to this method as the **AR benchmark**.

1. The surgeon places the HoloLens on her/his head and performs eye calibration to determine the pupil distance.
2. The application is started and the surgeon manually places a spatial anchor coordinate system next to the cadaver (see He et al. (2021) for details).
3. A 3D printed pointing device equipped with an AprilTag is tracked by the HoloLens. The surgeon digitizes the positions of the pedicle screw heads manually and confirms the precision of each acquired point after visual inspection of the overlay.
4. After all pedicle screw head positions were digitized, a virtual model of the target rod is presented to the surgeon and placed beside the bending bench.
5. The bending time is started when the lever of the bending bench is pushed down for the first time.
6. The surgeon bends the rod and compares the resulting implant to the virtual model.
7. The surgeon declares the rod ready for reduction into the anatomy.
8. If the rod does not yet meet clinical requirements, in-situ and ex-situ rebending maneuvers are required and counted.
9. The bending time is stopped after successful insertion of the rod into the pedicle screw heads.

The experimental protocol for **our** approach was defined as follows:

1. The surgeon places the HoloLens on her/his head and performs eye calibration to determine the pupil distance.
2. The application is started and the surgeon manually places a spatial anchor coordinate system next to the cadaver.
3. The surgeon is prompted to look at the anatomy and to change the head position by moving along the cadaver. The application stores incoming frames and infers bounding boxes using our stereo neural network. Rod calculation starts as soon as triangulation and clustering converged to a set of N_{screws} clusters. At this point, the surgeon can visually verify the found screw positions or restart the screw detection procedure in case of insufficient positional accuracy.
4. In the next step, the bending bench, which is rigidly attached to the table, is registered through an AprilTag. The surgeon can verify the registration by checking the overlay with the actual bending bench and either repeat registration or proceed with a voice command.
5. The surgeon fixes a forceps to the rod and aligns the real-world forceps with the overlay.
6. The bending procedure and the bending time are started. The surgeon can navigate through the bending steps using voice commands.
7. After the last bending step, an overlay of the desired final rod shape appears and the surgeon is allowed to perform final adaptations.
8. If the rod does not yet meet clinical requirements, in-situ and ex-situ rebending maneuvers are required and counted.
9. The bending time is stopped after successful insertion of the rod into the pedicle screw heads.

To avoid learning effects, the surgeons were invited twice with two-month intervals between trials. In the first experimental session, the surgeons were supported by the AR benchmark. In the second session, our approach was executed before the freehand benchmark. The cadavers which were assigned to the surgeons were swapped between sessions. It is worth noting that the surgeons were experienced in both benchmark methods, whereas it was their very first time using our navigation approach.

3.3.3. Statistical evaluation

Normal distribution of the data for each of the three navigation methods was tested with the Kolmogorov-Smirnov test (significance level $\alpha = 0.05$) for the bending time and the number of rebending attempts, respectively. Statistical differences between our proposed method and the benchmark approaches in terms of bending time and rebending maneuvers were analyzed using a two-sample t-test (significance level $\alpha = 0.05$).

4. Results

The results will be presented in three parts. Firstly, the mAP of the screw head bounding box detection is evaluated (Section 4.1) and the corresponding 3D accuracy is analyzed (Section 4.1.2). Secondly, results about our experimental validation on the transfer

function (mapping lever angle Θ to bending angle β) will be given. Lastly, the overall experimental evaluation of our navigation approach with respect to bending time and number of rebending attempts will be presented.

4.1. Pedicle screw head detection

The pedicle screw head detection evaluation is structured in two parts. Firstly, the performance of the network will be presented. Secondly, we will show the results of a 3D analysis to assess the average distance error of the triangulated and clustered points.

4.1.1. Network performance

The performance of the model was evaluated using 10-fold cross-validation. To this end, the mAP for each test fold was calculated using an intersection over union (IoU) threshold of 0.25 between predicted and ground-truth bounding box. Subsequently, the ten obtained mAP's were averaged to the final mAP estimate. Although an IoU-threshold of 0.25 was deemed sufficient by our surgeons to achieve clinical accuracy, an additional cross-validation experiment with a IoU of 0.5 was conducted to allow better comparability to other object detection algorithms. The evaluation of the 10-fold cross-validation resulted in an $mAP_{0.25}$ of $71.19 \pm 3.03\%$ ranging from 67.26% to 76.51%. When considering an IoU of 0.5 for true positives, the $mAP_{0.5}$ amounted to $32.63 \pm 2.03\%$ ranging from 28.83% to 35.88%.

Inference on an NVIDIA Quadro RTX 6000 took 27 ms for one stereo image pair resulting in a maximum possible throughput of 37fps which is well suited for real-time applications.

4.1.2. 3D analysis

The 3D positional accuracy of the pedicle screw centers was assessed by comparison to a CT ground truth (see Section 3.3.1). For all five runs, the stereo images and all relevant transformation matrices were recorded. The average positional 3D error (5 runs, 10 rods) was 5.43 mm. To estimate the error introduced by prediction inaccuracies, the predictions of the network were replaced by manual annotations of the same recorded images. After triangulation and clustering using the recorded transformations, a mean 3D distance error of 2.49 mm per rod was found. On average, our method took 49 s to digitize all screw heads compared to 67 s measured for the AR benchmark.

4.2. Transfer function

The transfer function mapping lever displacement angle Θ to observed bending angle β was obtained in a least square sense from the training data shown in Fig. 6. The following linear approximation was found:

$$\Theta = f(\beta) = 0.5653 \cdot \beta + 7.9836$$

To validate the acquired mapping, an additional 10 bends were performed where the target bending angles were set to $\beta = \{2^\circ, 5^\circ, 10^\circ, 15^\circ, 20^\circ, 25^\circ, 30^\circ, 35^\circ, 40^\circ, 45^\circ\}$ while the lever angle Θ was navigated. The RMSE of the linear approximation to unseen test data for β was 1.6° (shown in red in Fig. 6).

4.3. Evaluation of step-by-step navigation

With our proposed approach, the mean bending time amounted to $231\text{s} \pm 79$ s as compared to 476 s±360 s achieved by our AR benchmark and 348 s±192 s obtained by the freehand method. The fastest run took 155 s and the longest trial lasted 421 s using our approach, whereas the extreme values of the AR benchmark were 157 s and 1328 s. The fastest bending process of the freehand method was 174 s and the longest took 778 s.

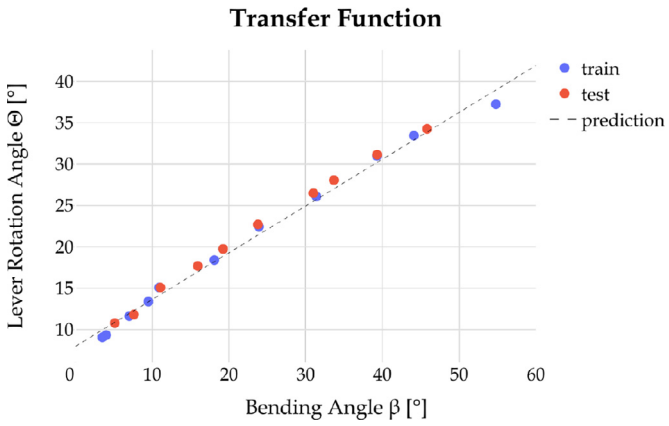


Fig. 6. Transfer function mapping lever rotation angle Θ to obtained bending angle in the rod β.

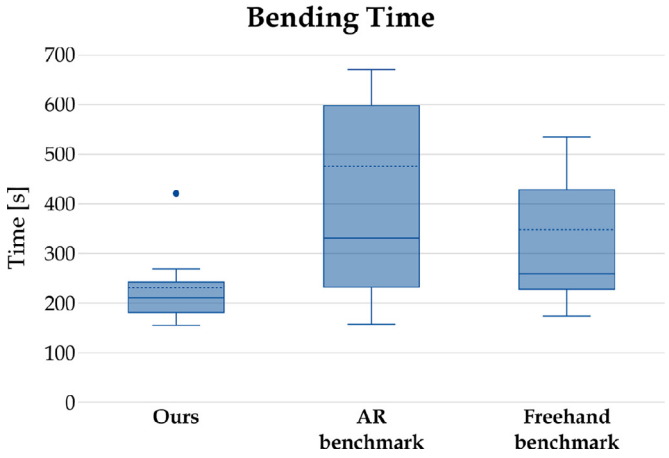


Fig. 7. Time required for bending and implantation of rod implants. The solid line in the box plot illustrates the median, whereas the dashed line marks the mean of the distribution.

The time distribution of the entire experiment is found in Fig. 7. Besides the bending time, the number of rebending maneuvers was recorded. On average, our approach needed an additional 0.6 ± 0.7 rebending maneuvers per rod. A total of four rods exhibited perfect fit without requiring additional rebendings. The maximum number of corrections recorded was 2. A mean of 1.1 ± 0.8 rebending maneuvers were observed for our AR benchmark ranging from 0 to 2. Contrary, the freehand method required 3.5 ± 3.0 rebending maneuvers varying from 0 to 8. An overview is given in Fig. 8. All presented results are summarized in Table 1.

The Kolmogorov-Smirnov test showed no evidence that our data was not normally distributed. Using our method, the number of required rebending maneuvers was significantly reduced with respect to the freehand method ($p = 0.04$), whereas we could not observe a difference compared to the AR benchmark ($p = 0.23$), or when comparing the two benchmark methods ($p = 0.08$). No significant decrease of the overall bending time was observed compared to the method of Wanivenhaus et al. (2019) ($p = 0.12$) or the freehand approach ($p = 0.17$), nor between the benchmark methods ($p = 0.43$).

5. Discussion

Improving the quality and safety of patient care and surgical outcome is of highest clinical relevance due to its enormous medical, social and economic impact (Kobayashi et al. (2018)). With approximately 200000 yearly performed elective fusion surgeries

Table 1
Overview of the results

	bending time [s]				rebendings [−]			
	mean	SD	min	max	mean	SD	min	max
Ours	231	79	155	421	0.6	0.7	0	2
AR benchmark	476	360	157	1328	1.1	0.8	0	2
Freehand benchmark	348	192	174	778	3.5	3.0	0	8

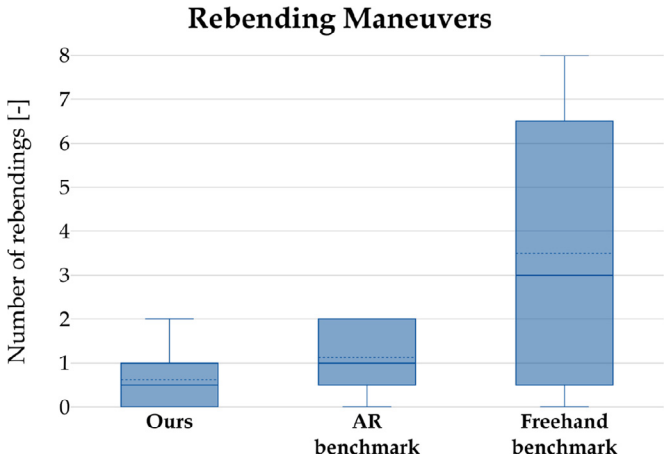


Fig. 8. The total number of rebending maneuvers required.

in the US alone (Martin et al., 2019), already a small decrease in perioperative complications would have a considerable impact on the quality of life of thousands of patients. Although computer-assisted surgery and surgical navigation could be a way to reduce surgical errors and thereby improve surgical outcomes, an overall impact on the treatment quality in orthopedic surgery has not been observed (Joskowicz and Hazan, 2016). In fact, only 5% of all orthopedic procedures are supported by surgical navigation (Joskowicz and Hazan, 2016). Reasons for the low acceptance are, among other things, the expensive hardware or the more complicated surgical workflow. Our approach tackles these shortcomings of current surgical navigation solutions. On the one hand, the elimination of markers greatly simplifies surgical navigation, reduces the risk of human errors and avoids line-of-sight problems (Wanivenhaus et al., 2019; Tohmeh et al., 2014). On the other hand, relying on the HoloLens provides a low-cost alternative to the specialized equipment required for other navigation approaches such as Bendini (Tohmeh et al., 2014).

We proposed a stereo neural network to directly estimate pedicle screw positions in a marker-less fashion. The combination of the union bounding box concept with a unified output representation allows us to obtain corresponding detections in stereo images. This network design enables training in an end-to-end fashion in contrast to our previous method (von Atzigen et al., 2020). We achieve a precision comparable to state-of-the-art object detection networks, but at the same time solve the challenging stereo correspondence problem. Tiny YOLO, for example, reports a $mAP_{0.5}$ of 33.1% while we obtain a $mAP_{0.5}$ of 32.63%. The measured decrease in the accuracy of our network from $mAP_{0.25}$ of 71.19% to $mAP_{0.5}$ of 33.1% appears to be an effect observable also in other bounding box based stereo networks (Konigshof et al., 2019). Instead of regressing a bounding box, the direct regression of screw center points may be an alternative to be investigated in the future. This would, however, come at the expense of losing versatility for future applications like the inclusion of other detectable objects such as the wound of the patient, surgical instruments, or the hands of surgeons. The screw detection performacnce resulted in an average error of 5.43 mm per rod. By replacing the detections of the network with manual annotations during accuracy evaluation (see Section 4.1.2), however, 2.49 mm of the average error could be attributed to tracking inaccuracies of the HoloLens. Network performance was further improved by data augmentation where the greatest benefits in the mAP were measured for geometric augmentations. This could be attributed to different people wearing the HoloLens for data acquisition. Vertical translation and flipping, for example, could simulate the inclusion of more viewpoints and scaling could be linked to varying heights of the subjects wearing the HoloLens. The smaller influence of arithmetic augmentations may be explained by the standardized recording environment, where the lighting conditions remained unchanged. Since we expect to encounter more dynamic lighting conditions in the OR environment such as bright spot-headlights from surgeons, shadow-casting by operating personnel, or reflections from excess blood, we believe that arithmetic augmentation could address the more challenging lighting conditions in these scenarios.

Besides data augmentation, the choice of a slower and more precise network could be an alternative to increase network performance. To this end, we have tested multiple network architectures including architectures requiring more computational power, and concluded that our proposed network architecture is a good trade-off between fast inference speeds and accuracy. A high frame rate is especially favorable since we require multiple forward passes to estimate the final screw positions.

All surgical navigation techniques discussed in Section 2.2 utilize AR as a visualization tool, but leave the execution of the task to human interpretation and dexterity. This is particularly surprising given literature indicating improved performance by providing step-by-step guidance in maintenance, assembly, or training (De Amicis et al., 2018; Webel et al., 2013). In this work, we propose a step-by-step instruction to the rod bending process which resulted in an average bending time of 231 s compared to 476 s and 348 s achieved by the AR and freehand benchmark, respectively. The same trend was observed for rebendings, where the average number of required rebendings with respect to the freehand method was reduced from 3.5 to 0.6 attempts. We associate the indications that the bending time could be reduced with a more intuitive guidance where no intermediate comparisons between the current and target rod shapes and no corrective bendings are necessary compared to our AR benchmark (Wanivenhaus et al., 2019). Previous studies have shown that not only the navigation method is of crucial importance but that also the visualization of the guidance plays a major role in the outcome of surgical procedures (Brendle et al., 2020). In this light, the smaller variability in the bending time and the number of rebending maneuvers indicates that our approach succeeds in providing clear, concise, and user-independent guidance.

However, our proposed pipeline has several limitations. Although the data for training the stereo neural network was obtained from different individuals with different professional backgrounds and varying HoloLens proficiency, a real OR environment may pose additional challenges. The depth of the wound in combination with the OR light and the surgeons' headlights may create more challenging illumination conditions for the screw detec-

tion. Even though previous studies have shown that the display and visualization capabilities of the HoloLens work well in the surgical environment (Dennler et al., 2021), further adoptions may be needed to fine-tune screw detection to the lighting conditions in the OR which will be addressed in future work. Furthermore, blood can cover parts of the screw and change its appearance in a way that makes detection less robust. However, the incorporation of training data from real-world OR environments or the inclusion of arithmetic augmentation techniques may help to mitigate these limitations. Another shortcoming is related to the hardware of the HoloLens, as the quality of the detections can be compromised by the poor quality of the grayscale stereo images. Additionally, we rely on the proprietary inside-out tracking of the HoloLens to merge the information from successive frames. Previous studies have investigated the influence of walking, sudden acceleration, sensor occlusion, and object insertion on hologram drift (Vassallo et al., 2017) which mainly depends on the head pose estimation. They reported an overall mean shift of 5.83 mm which could consequently introduce relevant deviations of the screw positions in the world coordinate system. Our triangulation and clustering routine reduces the influence of the HoloLens' proprietary SLAM and thus likely provides increased robustness. Despite the discussed challenges, AR offers significant advantages. On the one hand, the HoloLens merges sensing and visualization capabilities and thus supersedes setup, calibration, and data transfer efforts required when working with multiple devices. On the other hand, AR is a very versatile technology where algorithms can be adapted to other surgical disciplines using the same hardware compared to custom devices that need to be specifically manufactured for each type of intervention. Additionally, a recent clinical feasibility study (Dennler et al., 2021) showed that surgeons are satisfied with the performance of the HoloLens in terms of image quality, accuracy of virtual objects, and wearing comfort. Lastly, the advantages of our AR visualization being tailored to a specific bending bench comes with the disadvantage that the surgeon no longer has the freedom to choose which surgical tools to use. Adaptation of our navigation to other surgical tools, implant diameters, or materials, however, could be incorporated in future versions of our algorithm.

The commercial Bendini navigation system reports less residual force exerted on the screw-bone interface due to an improved rod shape (Tohmeh et al., 2014). In a future study, we aim at assessing this correlation for our approach quantitatively, as it could indicate a reduced risk of screw pull-out and hence revision surgery. Additionally, we target to analyze the rod quality with respect to surgeon proficiency. The decreased bending time in combination with fewer rebending maneuvers suggests an intuitive and user-independent navigation. Considering that the surgeons used our method for the first time, we believe that particularly novices could benefit from this kind of AR guidance.

6. Conclusion

A marker-less navigation approach to the rod bending process in spinal fusion surgery was presented which leverages the advantages of deep learning and AR. The key to our AR-based step-by-step navigation is the transformation of detected pedicle screw head positions to bending parameters which are eventually used to augment a surgical bending bench to guide the surgeon. We demonstrated that our proposed pipeline not only has the potential of reducing rod bending time but also significantly lowers the number of rebending maneuvers compared to the freehand benchmark technique. As a next milestone, we aim at accurately assessing forces that the rod implant exerts on the pedicle screws to better understand how the rod shape could be optimized. At a later stage, we would like to translate our method to the in-vivo treatment. We will further investigate in future work if a robotic agent could provide consistent and precise rod bending capabilities which would shorten surgery time and consequently make interventions cheaper.

Declaration of Competing Interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Prof. Dr. Mazda Farshad is shareholder and member of the board of directors of Incremed AG, a company developing mixed-reality applications. All other authors declare that they have no conflict of interest.

CRediT authorship contribution statement

Marco von Atzigen: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Data curation, Writing – original draft, Writing – review & editing, Visualization, Project administration. **Florentin Liebmann:** Software, Validation, Investigation. **Armando Hoch:** Conceptualization, Investigation. **José Miguel Spirig:** Conceptualization, Investigation. **Mazda Farshad:** Validation, Investigation, Resources, Funding acquisition. **Jess Snedeker:** Resources, Supervision. **Philipp Fürnstahl:** Conceptualization, Resources, Supervision, Funding acquisition.

Acknowledgments

This work is part of the SURGENT project under the umbrella of the Hochschulmedizin Zürich.

References

Abe, Y., Sato, S., Kato, K., Hyakumachi, T., Yanagibashi, Y., Ito, M., Abumi, K., 2013. A novel 3D guidance system using augmented reality for percutaneous vertebroplasty. Journal of Neurosurgery: Spine 19 (4), 492–501. doi:10.3171/2013.7.SPINE12917. https://pubmed.ncbi.nlm.nih.gov/23952323/

Ackermann, J., Liebmann, F., Hoch, A., Snedeker, J.G., Farshad, M., Rahm, S., Zingg, P.O., Fürnstahl, P., 2021. Augmented reality based surgical navigation of complex pelvic osteotomiesa feasibility study on cadavers. Applied Sciences (Switzerland) 11 (3), 1–19. doi:10.3390/app11031228.

Andress, S., Johnson, A., Unberath, M., Winkler, A., Yu, K., Fotouhi, J., Weidert, S., Osgood, G., Navab, N., 2018. On-the-fly augmented reality for orthopaedic surgery using a multi-modal fiducial. https://www.spiedigitallibrary.org/terms-of-use. 10.1117/1.jmi.5.2.021209.

von Atzigen, M., Liebmann, F., Hoch, A., Bauer, D.E., Snedeker, J.G., Farshad, M., Fürnstahl, P., 2020. Holoyolo: a proof-concept study for marker-less surgical navigation of spinal rod implants with augmented reality and on-device machine learning. The International Journal of Medical Robotics and Computer Assisted Surgery doi:10.1002/rcs.2184.

Barbanti-Brodano, G., Griffoni, C., Halme, J., Tedesco, G., Terzi, S., Bandiera, S., Ghermandi, R., Evangelisti, G., Girolami, M., Pipola, V., Gasbarrini, A., Falavigna, A., 2020. Spinal surgery complications: an unsolved problem-is the world health organization safety surgical checklist an useful tool to reduce them? European Spine Journal 29 (5), 927–936. doi:10.1007/s00586-019-06203-x.

Barry, P.J., Goldman, R.N., 1988. Recursive evaluation algorithm for a class of catmull-Rom splines. Computer Graphics (ACM) 22 (4), 199–204. doi:10.1145/378456.378511.

Baumberg, A., 2000. Reliable feature matching across widely separated views. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. IEEE, pp. 774–781. doi:10.1109/cvpr.2000.855899.

Bay, H., Tuytelaars, T., Van Gool, L., 2006. SURF: Speeded up robust features. In: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), Vol. 3951 LNCS, pp. 404–417. doi:10.1007/11744023_32.

Besl, P. J., McKay, N. D., 1992. Method for registration of 3-D shapes. In: https://doi.org/10.1117/12.57955.SPIE, pp. 586–606. https://www.spiedigitallibrary.org/conference-proceedings-of-spie/1611/0000/Method-for-registration-of-3-D-shapes/10.1117/12.57955.fullhttps://www.spiedigitallibrary.org/conference-proceedings-of-spie/1611/0000/Method-for-registration-of-3-D-shapes/10.1117/12.57955.short. 10.1117/12.57955.

Bo, L., Ren, X., Fox, D., 2014. Learning hierarchical sparse features for RGB-(D) object recognition. In: International Journal of Robotics Research. SAGE Publications Inc., pp. 581–599. doi:10.1177/0278364913514283.

Brachmann, E., Krull, A., Michel, F., Gumhold, S., Shotton, J., Rother, C., 2014. Learning 6D object pose estimation using 3D object coordinates. In: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). Springer Verlag, pp. 536–551. doi:10.1007/978-3-319-10605-2_35.

Brendle, C., Schütz, L., Esteban, J., Krieg, S.M., Eck, U., Navab, N., 2020. Can a Hand-Held Navigation Device Reduce Cognitive Load? A User-Centered Approach Evaluated by 18 Surgeons. In: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). Springer Science and Business Media Deutschland GmbH, pp. 399–408. doi:10.1007/978-3-030-59716-0_38.

Burström, G., Nachabe, R., Persson, O., Edström, E., Elmi Terander, A., 2019. Augmented and virtual reality instrument tracking for minimally invasive spine surgery: A Feasibility and accuracy study. Spine 44 (15), 1097–1104. doi:10.1097/BRS.0000000000003006. https://pubmed.ncbi.nlm.nih.gov/30830046/

Chen, X., Xu, L., Wang, Y., Wang, H., Wang, F., Zeng, X., Wang, Q., Egger, J., 2015. Development of a surgical navigation system based on augmented reality using an optical see-through head-mounted display. J Biomed Inform 55, 124–131. doi:10.1016/j.jbi.2015.04.003. https://pubmed.ncbi.nlm.nih.gov/25882923/

Cho, H.S., Park, M.S., Gupta, S., Han, I., Kim, H.S., Choi, H., Hong, J., 2018. Can augmented reality be helpful in pelvic bone cancer surgery? an in vitro study. Clin. Orthop. Relat. Res. 476 (9), 1719–1725. doi:10.1007/s11999.0000000000000233. https://pubmed.ncbi.nlm.nih.gov/30794209/

Dalstrom, D.J., Venkatarayappa, I., Manternach, A.L., Palcic, M.S., Heyse, B.A., Prayson, M.J., 2008. Time-dependent contamination of opened sterile operating-room trays. Journal of Bone and Joint Surgery - Series A 90 (5), 1022–1025. doi:10.2106/JBJS.G.00689. https://journals.lww.com/jbjsjournal/Fulltext/2008/05000/Time_Dependent_Contamination_of_Opened_Sterile.11.aspx

De Amicis, R., Ceruti, A., Francia, D., Frizziero, L., Simões, B., 2018. Augmented reality for virtual user manual. Int. J. Interact. Des. Manuf. 12 (2), 689–697. doi:10.1007/s12008-017-0451-7.

Deib, G., Johnson, A., Unberath, M., Yu, K., Andress, S., Qian, L., Osgood, G., Navab, N., Hui, F., Gailloud, P., 2018. Image guided percutaneous spine procedures using an optical see-through head mounted display: proof of concept and rationale. J Neurointerv Surg 10 (12), 1187–1191. doi:10.1136/neurintsurg-2017-013649. https://pubmed.ncbi.nlm.nih.gov/29848559/

Dennler, C., Bauer, D.E., Scheibler, A.G., Spirig, J., Götschi, T., Fürnstahl, P., Farshad, M., 2021. Augmented reality in the operating room: a clinical feasibility study. BMC Musculoskelet Disord 22 (1), 1–9. doi:10.1186/s12891-021-04339-w.

Deriche, R., Zhang, Z., Luong, Q.T., Faugeras, O., 1994. Robust recovery of the epipolar geometry for an uncalibrated stereo rig. In: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), Vol. 800 LNCS, pp. 567–576. doi:10.1007/3-540-57956-7_64.

Du, X., Kurmann, T., Chang, P.L., Allan, M., Ourselin, S., Sznitman, R., Kelly, J.D., Stoyanov, D., 2018. Articulated multi-instrument 2-d pose estimation using fully convolutional networks. IEEE Trans Med Imaging 37 (5), 1276–1287. doi:10.1109/TMI.2017.2787672.

Durrant-Whyte, H., Bailey, T., 2006. Simultaneous localization and mapping: Part I. IEEE Rob. Autom. Mag. 13 (2), 99–108. doi:10.1109/MRA.2006.1638022.

Elmi-Terander, A., Burström, G., Nachabe, R., Skulason, H., Pedersen, K., Fagerlund, M., Ståhl, F., Charalampidis, A., Söderman, M., Holmin, S., Babic, D., Jenniskens, I., Edström, E., Gerdhem, P., 2019. Pedicle screw placement using augmented reality surgical navigation with intraoperative 3D imaging: A First in-Human prospective cohort study. Spine 44 (7), 517–525. doi:10.1097/BRS.0000000000002876. www.spinejournal.com

Fischler, M.A., Bolles, R.C., 1981. Random sample consensus: A Paradigm for model fitting with applications to image analysis and automated cartography. Commun ACM 24 (6), 381–395. doi:10.1145/358669.358692.

He, W., Xi, M., Gardner, H., Swift, B., Adcock, M., 2021. Spatial anchor based indoor asset tracking. IEEE Virtual Reality and 3D User Interfaces (VR).

Hein, J., Seibold, M., Bogo, F., Farshad, M., Pollefeys, M., Fürnstahl, P., Navab, N., 2021. Towards markerless surgical tool and hand pose estimation. Int J Comput Assist Radiol Surg 1–10. doi:10.1007/s11548-021-02369-2.

Hoch, A., Liebmann, F., Carrillo, F., Farshad, M., Rahm, S., Zingg, P.O., Fürnstahl, P., 2021. Augmented reality based surgical navigation of the periacetabular osteotomy of ganz - a pilot cadaveric study. In: Mechanisms and Machine Science, Vol. 93. Springer Science and Business Media B.V., pp. 192–201. doi:10.1007/978-3-030-58104-6_22.

Hu, Y., Fua, P., Wang, W., Salzmann, M., 2020. Single-Stage 6D Object Pose Estimation. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 2927–2936. doi:10.1109/CVPR42600.2020.00300.

Hu, Y., Hugonot, J., Fua, P., Salzmann, M., 2018. Segmentation-driven 6D object pose estimation.

Joskowicz, L., Hazan, E. J., 2016. Computer Aided Orthopaedic Surgery: Incremental shift or paradigm change?10.1016/j.media.2016.06.036.

Jud, L., Fotouhi, J., Andronic, O., Aichmair, A., Osgood, G., Navab, N., Farshad, M., 2020. Applicability of augmented reality in orthopedic surgery - A systematic review. BMC Musculoskelet Disord 21 (1), 103. doi:10.1186/s12891-020-3110-2. http://www.ncbi.nlm.nih.gov/pubmed/32061248 http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC7023780

Kadkhodamohammadi, A., 2016. 3D detection and pose estimation of medical staff in operating rooms using RGB-D images. Technical Report. https://tel.archives-ouvertes.fr/tel-01553825

Kehl, W., Milletari, F., Tombari, F., Ilic, S., Navab, N., 2016. Deep learning of local RGB-D patches for 3D object detection and 6D pose estimation. In: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), Vol. 9907 LNCS, pp. 205–220. doi:10.1007/978-3-319-46487-9_13.

Kobayashi, K., Ando, K., Nishida, Y., Ishiguro, N., Imagama, S., 2018. Epidemiological trends in spine surgery over 10 years in a multicenter database. European Spine Journal 27 (8), 1698–1703. doi:10.1007/s00586-018-5513-4.

Konigshof, H., Salscheider, N.O., Stiller, C., 2019. Realtime 3D Object Detection for Automated Driving Using Stereo Vision and Semantic Information. In: 2019 IEEE Intelligent Transportation Systems Conference, ITSC 2019, pp. 1405–1410. doi:10.1109/ITSC.2019.8917330.

Kurmann, T., Marquez Neila, P., Du, X., Fua, P., Stoyanov, D., Wolf, S., Sznitman, R., 2017. Simultaneous recognition and pose estimation of instruments in minimally invasive surgery. In: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), Vol. 10434 LNCS, pp. 505–513. doi:10.1007/978-3-319-66185-8_57.

Laine, T., Schlenzka, D., Mäkitalo, K., Tallroth, K., Nolte, L.P., Visarius, H., 1997. Improved accuracy of pedicle screw insertion with computer-assisted surgery: a prospective clinical trial of 30 patients. Spine 22 (11), 1254–1258. doi:10.1097/00007632-199706010-00018. https://pubmed.ncbi.nlm.nih.gov/9201865/

Laverdière, C., Corban, J., Khoury, J., Ge, S.M., Schupbach, J., Harvey, E.J., Reindl, R., Martineau, P.A., 2019. Augmented reality in orthopaedics: a systematic review and a window on future possibilities. Bone and Joint Journal 101-B (12), 1479–1488. doi:10.1302/0301-620X.101B12.BJJ-2019-0315.R1. https://pubmed.ncbi.nlm.nih.gov/31786992/

Lepetit, V., Moreno-Noguer, F., Fua, P., 2009. EPnP: an accurate o(n) solution to the PnP problem. Int J Comput Vis 81 (2), 155–166. doi:10.1007/s11263-008-0152-6. http://cvlab.epfl.ch/software/EPnP/

Li, P., Chen, X., Shen, S., 2019. Stereo R-CNN based 3D object detection for autonomous driving. Technical Report doi:10.1109/CVPR.2019.00783. https://github.com/HKUST-Aerial-Robotics/Stereo-RCNN

Liebmann, F., Roner, S., von Atzigen, M., Scaramuzza, D., Sutter, R., Snedeker, J., Farshad, M., Fürnstahl, P., 2019. Pedicle screw navigation using surface digitization on the microsoft hololens. Int J Comput Assist Radiol Surg 14 (7), 1157–1165. doi:10.1007/s11548-019-01973-7.

Liu, B., Li, Y., Zhang, S., Ye, X., 2017. Healthy human sitting posture estimation in RGB-D scenes using object context. Multimed Tools Appl 76 (8), 10721–10739. doi:10.1007/s11042-015-3189-x. http://www.openni.ru/

Liu, H., Auvinet, E., Giles, J., Rodriguez y Baena, F., 2018. Augmented reality based navigation for computer assisted hip resurfacing: A Proof of concept study. Ann Biomed Eng 46 (10), 1595–1605. doi:10.1007/s10439-018-2055-1.

Lowe, D.G., 2004. Distinctive image features from scale-invariant keypoints. Int J Comput Vis 60 (2), 91–110. doi:10.1023/B:VISI.0000029664.99615.94.

Martin, B.I., Mirza, S.K., Spina, N., Spiker, W.R., Lawrence, B., Brodke, D.S., 2019. Trends in lumbar fusion procedure rates and associated hospital costs for degenerative spinal diseases in the united states, 2004 to 2015. Spine 44 (5), 369–376. doi:10.1097/BRS.0000000000002822.

Mavrogenis, A.F., Savvidou, O.D., Mimidis, G., Papanastasiou, J., Koulalis, D., Demertzis, N., Papagelopoulos, P.J., 2013. Computer-assisted navigation in orthopedic surgery. Orthopedics 36 (8), 631–642. doi:10.3928/01477447-20130724-10.

Menekse, G., Kuscu, F., Suntur, B.M., Gezercan, Y., Ates, T., Ozsoy, K.M., Okten, A.I., 2015. Evaluation of the time-dependent contamination of spinal implants. Spine 40 (16), 1247–1251. doi:10.1097/BRS.0000000000000944. https://journals.lww.com/spinejournal/Fulltext/2015/08150/Evaluation_of_the_Time_dependent_Contamination_of.3.aspx

Merloz, P., Tonetti, J., Cinquin, P., Lavallée, S., Troccaz, J., Pittet, L., 1998. Computer assisted pedicle screw placement. Chirurgie 123 (5), 482–490. doi:10.1016/s0001-4001(99)80077-4. https://pubmed.ncbi.nlm.nih.gov/9882919/

Nasser, R., Yadla, S., Maltenfort, M. G., Harrop, J. S., Anderson, G., Vaccaro, A. R., Sharan, A. D., Ratliff, J. K., 2010. Complications in spine surgery a review. https://thejns.org/spine/view/journals/j-neurosurg-spine/13/2/article-p144.xml. 10.3171/2010.3.SPINE09369.

Nguyen, N.Q., Cardinell, J., Ramjist, J., Dobashi, Y., Androutsos, D., Yang, V.X.D., 2019. Augmented reality systems for improved operating room workflow. Neurosurgery 66 (Supplement_1). doi:10.1093/neuros/nyz310_414. https://academic.oup.com/neurosurgery/article/66/Supplement_1/nyz310_414/5551922

Ni, Z.L., Bian, G.B., Xie, X.L., Hou, Z.G., Zhou, X.H., Zhou, Y.J., 2019. RASNet: Segmentation for Tracking Surgical Instruments in Surgical Videos Using Refined Attention Segmentation Network. In: Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS, pp. 5735–5738. doi:10.1109/EMBC.2019.8856495.

Nottmeier, E.W., Crosby, T.L., 2007. Timing of paired points and surface matching registration in three-dimensional (3D) image-guided spinal surgery. Journal of Spinal Disorders and Techniques 20 (4), 268–270. doi:10.1097/01.bsd.0000211282.06519.ab.

Parchami, M., Cadeddu, J.A., Mariottini, G.L., 2014. Endoscopic stereo reconstruction: A comparative study. In: 2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBC 2014. Institute of Electrical and Electronics Engineers Inc., pp. 2440–2443. doi:10.1109/EMBC.2014.6944115.

Pavlakos, G., Zhou, X., Chan, A., Derpanis, K.G., Daniilidis, K., 2017. 6-DoF object pose from semantic keypoints. In: Proceedings - IEEE International Conference on Robotics and Automation, pp. 2011–2018. doi:10.1109/ICRA.2017.7989233.

Peng, S., Liu, Y., Huang, Q., Zhou, X., Bao, H., 2019. PVNET: Pixel-wise voting network for 6dof pose estimation. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Vol. 2019-June, pp. 4556–4565. doi:10.1109/CVPR.2019.00469. https://zju-3dv.github.io/pvnet/.

Pritchett, P., Zisserman, A., 1998. Wide baseline stereo matching. In: Proceedings of the IEEE International Conference on Computer Vision. IEEE, pp. 754–760. doi:10.1109/iccv.1998.710802.

Probst, T., Maninis, K.K., Chhatkuli, A., Ourak, M., Poorten, E.V., Van Gool, L., 2018. Automatic tool landmark detection for stereo vision in robot-Assisted retinal surgery. IEEE Rob. Autom. Lett. 3 (1), 612–619. doi:10.1109/LRA.2017.2778020. https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=8120159

Rad, M., Lepetit, V., 2017. BB8: A Scalable, Accurate, Robust to Partial Occlusion Method for Predicting the 3D Poses of Challenging Objects without Using Depth. In: Proceedings of the IEEE International Conference on Computer Vision, Vol. 2017-Octob, pp. 3848–3856. doi:10.1109/ICCV.2017.413.

Redmon, J., Farhadi, A., 2018. YOLOv3: An Incremental Improvementhttps://pjreddie.com/yolo/. http://arxiv.org/abs/1804.02767. 10.1109/CVPR.2017.690.

Richter, M., Cakir, B., Schmidt, R., 2005. Cervical pedicle screws: conventional versus computer-assisted placement of cannulated screws. Spine 30 (20), 2280–2287. doi:10.1097/01.brs.0000182275.31425.cd.

Salah, Z., Preim, B., Elolf, E., Franke, J., Rose, G., 2011. Improved navigated spine surgery utilizing augmented reality visualization. In: Informatik aktuell. Springer, Berlin, Heidelberg, pp. 319–323. doi:10.1007/978-3-642-19335-4_66.

Scharstein, D., Szeliski, R., Zabih, R., 2001. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. In: Proceedings - IEEE Workshop on Stereo and Multi-Baseline Vision, SMBV 2001, Vol. 47, pp. 131–140. doi:10.1109/SMBV.2001.988771. www.middlebury.edu/stereo.

Schlenzka, D., Laine, T., Lund, T., 2000. Computer-assisted spine surgery. European Spine Journal 9 (SUPPL.1), S057–S064. doi:10.1007/pl00010023.

Schwarz, M., Schulz, H., Behnke, S., 2015. RGB-D object recognition and pose estimation based on pre-trained convolutional neural network features. In: Proceedings - IEEE International Conference on Robotics and Automation. Institute of Electrical and Electronics Engineers Inc., pp. 1329–1335. doi:10.1109/ICRA.2015.7139363.

Shvets, A.A., Rakhlin, A., Kalinin, A.A., Iglovikov, V.I., 2019. Automatic Instrument Segmentation in Robot-Assisted Surgery using Deep Learning. In: Proceedings - 17th IEEE International Conference on Machine Learning and Applications, ICMLA 2018, pp. 624–628. doi:10.1109/ICMLA.2018.00100.

Sorko, S.R., Brunnhofer, M., 2019. Potentials of Augmented Reality in Training. In: Procedia Manufacturing. Elsevier B.V., pp. 85–90. doi:10.1016/j.promfg.2019.03.014.

Sridhar, S., Mueller, F., Zollhöfer, M., Casas, D., Oulasvirta, A., Theobalt, C., 2016. Real-time joint tracking of a hand manipulating an object from RGB-D input. In: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), Vol. 9906 LNCS, pp. 294–310. doi:10.1007/978-3-319-46475-6_19.

Tan, D. J., Navab, N., Tombari, F., 2017. 6D Object Pose Estimation with Depth Images: A Seamless Approach for Robotic Interaction and Augmented Reality.

Tekin, B., Sinha, S.N., Fua, P., 2018. Real-Time Seamless Single Shot 6D Object Pose Prediction. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 292–301. doi:10.1109/CVPR.2018.00038.

Tohmeh, A., Isaacs, R.E., Dooley, Z.A., Turner, A.W., 2014. Long construct pedicle screw reduction and residual forces are decreased using a computer-Assisted spinal rod bending system. The Spine Journal 14 (11), S143–S144. doi:10.1016/j.spinee.2014.08.348.

Uzun, E., Misir, A., Ozcamdalli, M., Kizkapan, E.E., Cirakli, A., Calgin, M.K., 2020. Time-dependent surgical instrument contamination begins earlier in the uncovered table than in the covered table. Knee Surgery, Sports Traumatology, Arthroscopy 28 (6), 1774–1779. doi:10.1007/s00167-019-05607-y.

Vassallo, R., Rankin, A., Chen, E.C.S., Peters, T.M., 2017. Hologram stability evaluation for Microsoft HoloLens. In: Medical Imaging 2017: Image Perception, Observer Performance, and Technology Assessment. SPIE, p. 1013614. doi:10.1117/12.2255831.

Wang, C., Guo, X., 2017. Feature-based RGB-D camera pose optimization for real-time 3D reconstruction. Computational Visual Media 3 (2), 95–106. doi:10.1007/s41095-016-0072-2.

Wang, W., Wang, F., Song, W., Su, S., 2020. Application of augmented reality (AR) technologies in inhouse logistics. In: E3S Web of Conferences. EDP Sciences doi:10.1051/e3sconf/202014502018.

Wanivenhaus, F., Neuhaus, C., Liebmann, F., Roner, S., Spirig, J.M., Farshad, M., 2019. Augmented reality-assisted rod bending in spinal surgery. Spine Journal 19 (10), 1687–1689. doi:10.1016/j.spinee.2019.06.019.

Webel, S., Bockholt, U., Engelke, T., Gavish, N., Olbrich, M., Preusche, C., 2013. An augmented reality training platform for assembly and maintenance skills. Rob Auton Syst 61 (4), 398–403. doi:10.1016/j.robot.2012.09.013.

Westerfield, G., Mitrovic, A., Billinghurst, M., 2015. Intelligent augmented reality training for motherboard assembly. Int. J. Artif. Intell. Educ. 25 (1), 157–172. doi:10.1007/s40593-014-0032-x.

Whelan, T., Johannsson, H., Kaess, M., Leonard, J.J., McDonald, J., 2013. Robust real-time visual odometry for dense RGB-D mapping. In: Proceedings - IEEE International Conference on Robotics and Automation, pp. 5724–5731. doi:10.1109/ICRA.2013.6631400.

Wu, Q., Xu, G., Zhang, S., Li, Y., Wei, F., 2020. Human 3D pose estimation in a lying position by RGB-D images for medical diagnosis and rehabilitation. In: Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS. Institute of Electrical and Electronics Engineers Inc., pp. 5802–5805. doi:10.1109/EMBC44109.2020.9176407.

Xiang, Y., Schmidt, T., Narayanan, V., Fox, D., 2017. PoseCNN: A convolutional neural network for 6D object pose estimation in cluttered scenes. https://rse-lab.cs.washington.edu/projects/posecnn/.10.15607/rss.2018.xiv.019.

Xie, H., Yao, H., Zhou, S., Zhang, S., Sun, X., Sun, W., 2019. Toward 3D object reconstruction from stereo images. https://www.blender.org.

Zeng, A., Song, S., Nießner, M., Fisher, M., Xiao, J., Funkhouser, T., 2017. 3DMatch: Learning local geometric descriptors from RGB-D reconstructions. In: Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, Vol. 2017-Janua, pp. 199–208. doi:10.1109/CVPR.2017.29. http://3dmatch.cs.princeton.eduhttp//3dmatch.cs.princeton.edu.

Zhang, H., Cao, Q., 2017. Texture-less object detection and 6D pose estimation in RGB-D images. Rob Auton Syst 95, 64–79. doi:10.1016/j.robot.2017.06.003.

Zhang, Z., Deriche, R., Faugeras, O., Luong, Q.T., 1995. A robust technique for matching two uncalibrated images through the recovery of the unknown epipolar geometry. Artif Intell 78 (1–2), 87–119. doi:10.1016/0004-3702(95)00022-4.

Zhu, Z., Branzoi, V., Wolverton, M., Murray, G., Vitovitch, N., Yarnall, L., Acharya, G., Samarasekera, S., Kumar, R., 2014. AR-mentor: Augmented reality based mentoring system. In: ISMAR 2014 - IEEE International Symposium on Mixed and Augmented Reality - Science and Technology 2014, Proceedings. Institute of Electrical and Electronics Engineers Inc., pp. 17–22. doi:10.1109/ISMAR.2014.6948404.

Zia, S., Yüksel, B., Yüret, D., Yemez, Y., 2017. RGB-D object recognition using deep convolutional neural networks. In: Proceedings - 2017 IEEE International Conference on Computer Vision Workshops, ICCVW 2017, Vol. 2018-Janua, pp. 887–894. doi:10.1109/ICCVW.2017.109.