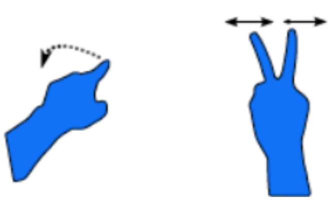
ISSN 2830-9820

Visual Computing MAGAZiNE

Vol. 3, Issue 1
2025

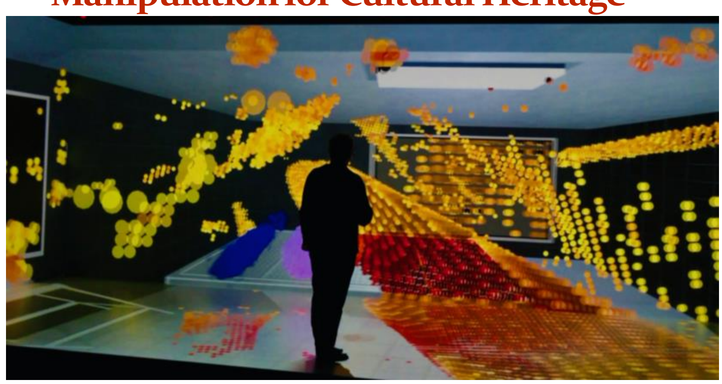
The Metaverse and Immersive Technologies

Deep Learning-Based Hand Pose Estimation: A Novel ResUnet Framework for Occlusion Handling in AR/VR Application



A Systematic Procedure for Comparing Template-based Gesture Recognizers

Immersive Volumetric Point Cloud Manipulation for Cultural Heritage



MACHINE AND DEEP LEARNING TECHNIQUES FOR HISTOLOGICAL AND ENDOSCOPIC IMAGES

Visual Computing Magazine

The Preface

When Professor Slimane Larabi asked me to be editor-in-chief of this issue of the magazine, I responded favorably almost instantly. I did not realize how difficult it would be to convince authors to publish in a magazine that is not referenced by Scopus or any other type of referencing. It was then necessary to explain to potential authors the quality of this magazine and its importance in bringing the world of research closer to our graduating students. I graduated from USTHB in the 1980s and I would have liked to have access to magazines of this quality. When choosing my final year project topic at the end of my engineering program, I decided with my partner to work on image processing, which was not taught in our curriculum. We nevertheless undertook this work with enthusiasm, while suffering from the lack of references and access to international bibliography. Today, fortunately, our students no longer have this problem.

Secondly, we had to choose a general, current theme to highlight new technologies in the field of image and visual arts. The metaverse and immersive technologies emerged naturally. The first article addresses these themes and introduces them so that the reader can understand the ins and outs of the Metaverse and Immersive Technologies, the concepts, the different approaches, the challenges, and the socio-economic impacts. Artificial intelligence dominates the media scene, and this is logical, but I am convinced that the Metaverse will follow very soon. These two are complementary, and several articles in this issue deal with AI in relation to images.
This issue then includes an article on the application of virtual reality to the field of archaeological heritage. Algeria is full of historical remains, and we must exploit new technologies for their preservation and restoration, and to make them accessible and visitable virtually.

The Metaverse and virtual and augmented reality require the most natural human interaction. The third and fourth articles address gestural interaction, which allows the user to move around the virtual scene and manipulate virtual objects.
The final article discusses machine and deep learning techniques for medical imaging. These same techniques are and will be used to generate 3D scenes needed to build virtual environments in the Metaverse.

My dream is for Algeria to position itself as an African leader in the Metaverse and immersive technologies, contributing significantly to addressing national priorities such as citizen health, food and water security, and energy security.

I would like to sincerely thank Professor Slimane Larabi for giving me the opportunity to host this issue of my home university, and I wish everyone a beneficial read.

The Guess Editor, Prof. Rachid GHERBI
University Paris Saclay, France

Visual Computing: *A quarterly magazine. Volume 3, Issue 1, 2025*
ISSN: **2830-9820**
Chief Editor: *Prof. Slimane LARABI*

Visual Computing Magazine

The Metaverse and Immersive Technologies

Concepts, Approaches, Challenges and Socio-Economical Impacts

Rachid Gherbi, Computer Science Department, Université Paris Saclay, France

Metaverse Definition and Concept

The term "Metaverse" refers to a virtual reality space where users can interact with a computer-generated environment and other users in real time. These users could just intercte with each others, but could perform common tasks in order to achieve a mission. This mission could be gaming, construction, working, communication, shopping, etc.

Metaverse is a word composed of two ones. The first one, Meta is a prefix that comes from the Greek μετά (meta) (after, beyond, with). It expresses at the same time reflection, change, succession, the fact of going beyond, beside, between or withis. The second word is Verse that means Univers, world, environment, space, etc.

The Metaverse is a concept that originated in science fiction (ref) but has gained popularity and attention in recent decade. It is often depicted as a fully immersive and interconnected virtual world that goes beyond traditional two-dimensional screens. It aims to provide a sense of presence and immersive social interaction in a digital environment that mimics the real world, creates entirely new and fantastical spaces, or mixes real and artificial univers.

The concept of the Metaverse encompasses various elements involving in one hand immersive technologies such as virtual reality (VR), augmented reality (AR), mixed or extended realities (MR&XR), and on the other hand technologies such as artificial intelligence (AI) and blockchain technology. Both technologies work together to create an interactive and immersive shared virtual space that can be accessed through different devices, such as VR and MR headsets and glasses, smartphones, or computers.

Visual Computing Magazine

The Metaverse and Immersive Technologies

Concepts, Approaches, Challenges and Socio-Economical Impacts

Rachid Gherbi, Computer Science Department, Université Paris Saclay, France

In the Metaverse, users can create avatars or digital representations of themselves to navigate and interact with the virtual world. They can explore virtual environments, engage in social activities, communicate with others through voice or text chat, participate in games and events, buy and sell virtual goods and assets, and even conduct business transactions.

The vision of the Metaverse goes beyond simple entertainment and gaming. It is envisioned as a platform that can facilitate various activities, such as education and training, healthcare, oil & gas and other industries, communication, art, commerce, travelling and tourism, and more. It has the potential to revolutionize how we interact, collaborate, and experience digital content.

It's important to note that the concept of the Metaverse is still evolving and the actual implementation and realization of a fully functional Metaverse is yet to be achieved. However, various technology companies, including big players like Facebook, Apple and chineese companies, are investing in research and development to make the Metaverse a reality in the closed future.

Challenges of the Metaverse

While the concept of the Metaverse holds immense potential, there are several challenges that need to be addressed for its successful implementation. Here are some of the key challenges:

1. Technical Infrastructure: Building a fully functional Metaverse requires robust technical infrastructure to support the massive amount of data, interactions, and real-time rendering involved. This includes high-speed internet connections, powerful servers, low-latency networks, and efficient data processing capabilities. Scaling these infrastructure components to accommodate millions or even billions of users simultaneously poses a significant challenge.

The Metaverse and Immersive Technologies

Concepts, Approaches, Challenges and Socio-Economical Impacts

Rachid Gherbi, Computer Science Department, Université Paris Saclay, France

Figure 1 : chalenges of metaverse
(https://www.revinfotech.com/blog/development-of-the-metaverse-challenges-and-remedies/)

2. Interoperability and Standards: For a true Metaverse experience, different platforms, devices, and applications need to seamlessly interact with each other. However, achieving interoperability and establishing common standards across diverse technologies, software frameworks, and hardware platforms is a complex task. Without interoperability, the Metaverse could become fragmented, limiting user experiences and hindering widespread adoption.

The Metaverse and Immersive Technologies

Concepts, Approaches, Challenges and Socio-Economical Impacts

Rachid Gherbi, Computer Science Department, Université Paris Saclay, France

3. Privacy and Security: As the Metaverse relies heavily on personal data, privacy and security become critical concerns. Protecting user information, preventing unauthorized access, and ensuring data integrity are paramount to build trust and confidence in the Metaverse. Robust privacy policies, secure authentication mechanisms, and effective data encryption techniques will be essential to address these challenges.

4. Content Creation and Curation: A thriving Metaverse requires a constant supply of engaging and diverse content. However, creating and curating vast amounts of high-quality content can be daunting. Balancing user-generated content with professionally produced content, managing intellectual property rights, and moderating user-generated interactions to maintain a safe and inclusive environment are all significant challenges.

5. Accessibility and Inclusivity: Ensuring equal access and inclusivity for all individuals, including those with disabilities, across different devices, networks, and socioeconomic backgrounds is crucial. Overcoming barriers such as affordability, connectivity limitations, and accessibility requirements will be essential to prevent a digital divide and ensure broad participation in the Metaverse.

6. Ethical and Social Implications: The Metaverse raises important ethical and social considerations. Issues such as digital addiction, loss of physical human connection, virtual identity manipulation, and the impact on mental health need to be carefully addressed. Clear guidelines and ethical frameworks should be established to govern the use and impact of the Metaverse on society.

Addressing these challenges requires collaboration among technology companies, policymakers, developers, content creators, and users to ensure the development of an inclusive, secure, and sustainable Metaverse that benefits society as a whole.

The Metaverse and Immersive Technologies

Concepts, Approaches, Challenges and Socio-Economical Impacts

Rachid Gherbi, Computer Science Department, Université Paris Saclay, France

Applications and Socio-Economical impacts of the Metaverse

The potential applications of the Metaverse span a wide range of industries and activities. Here are some examples of how the Metaverse could be applied:

1. Social Networking and Communication: The Metaverse can provide a rich and immersive platform for social interaction. Users can connect with friends, family, and colleagues in virtual environments, attend virtual events, and engage in real-time communication through voice, video, and chat.

2. Gaming and Entertainment: The gaming industry stands to benefit greatly from the Metaverse. It can enhance multiplayer gaming experiences, enabling players to interact in shared virtual worlds, participate in massive online tournaments, and explore immersive story-driven narratives.

Figure 2: Social Networking and Communication xithin the Metaverse; (https://www.metaengine.gg/blog/how-is-the-metaverse-disrupting-traditional-social-interactions)

The Metaverse and Immersive Technologies

Concepts, Approaches, Challenges and Socio-Economical Impacts

Rachid Gherbi, Computer Science Department, Université Paris Saclay, France

3- Education and Training: The Metaverse can revolutionize education and training by creating interactive and immersive learning environments. Students can attend virtual classrooms, explore historical simulations, conduct scientific experiments, and engage in collaborative projects with peers from around the world.

Figure 3: gaming in the Metaverse (https://www.xrtoday.com/virtual-reality/gaming-in-the-metaverse-the-next-frontier/)

4. Virtual Commerce: The Metaverse has the potential to transform e-commerce by providing a virtual marketplace where users can buy and sell virtual goods, digital assets, and even physical products. Virtual showrooms, customized virtual stores, and immersive shopping experiences can enhance the way people shop and trade online.

Figure 4: education within the Metaverse (https://99infosystems.com/metaverse-in-education-new-learning-paradigm/)

The Metaverse and Immersive Technologies

Concepts, Approaches, Challenges and Socio-Economical Impacts

Rachid Gherbi, Computer Science Department, Université Paris Saclay, France

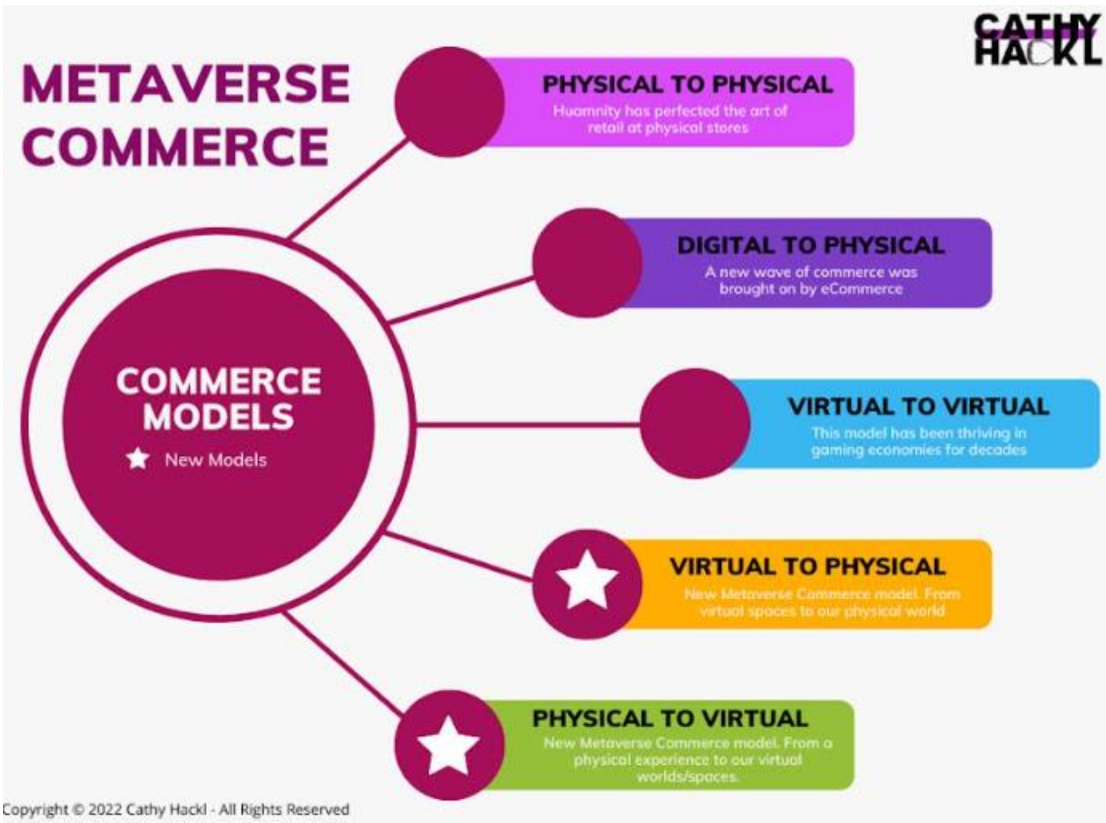


Figure 5: commerce in the Metaverse (https://www.forbes.com/sites/cathyhackl/2022/07/05/metaverse-commerce-understanding-the-new-virtual-to-physical-and-physical-to-virtual-commerce-models/)

5- Virtual Workplaces: With remote work becoming more prevalent, the Metaverse can offer virtual office spaces where employees can collaborate, attend virtual meetings, and work together in a shared digital environment. It can provide a sense of presence and make remote work more engaging and interactive.

The Metaverse and Immersive Technologies

Concepts, Approaches, Challenges and Socio-Economical Impacts

Rachid Gherbi, Computer Science Department, Université Paris Saclay, France

Figure 6: workplaces within the Metaverse
(https://www.servcorp.com.au/en/blog/technology/the-future-of-work-what-an-office-in-the-metaverse-could-look-like/)

6- Healthcare and Therapy: The Metaverse can be utilized for telemedicine and virtual healthcare services, allowing doctors to diagnose and treat patients remotely. It can also be used for therapy and rehabilitation, providing immersive environments for mental health support and physical rehabilitation programs.

7- Art, Music, and Creativity: The Metaverse can serve as a platform for artists, musicians, and creators to showcase their work, collaborate with others, and engage with audiences in new and innovative ways. Virtual galleries, concerts, and interactive art installations can bring immersive creative experiences to a global audience.

The Metaverse and Immersive Technologies

Concepts, Approaches, Challenges and Socio-Economical Impacts

Rachid Gherbi, Computer Science Department, Université Paris Saclay, France

Figure 7: Healthcare and Metaverse
(https://healthcarebusinessclub.com/articles/healthcare-provider/technology/metaverse-in-healthcare/)

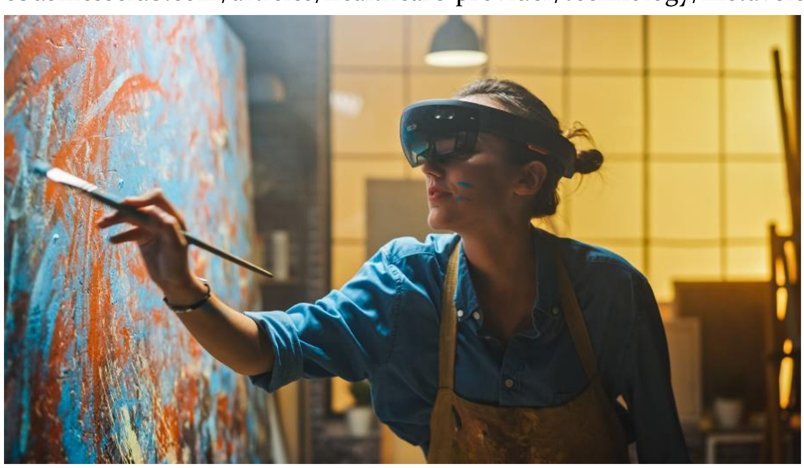


Figure 8: Art and Metaverse
(https://www.stambol.com/2021/05/31/digital-art-nfts-are-redefining-real/)

8- Virtual Tourism and Exploration: The Metaverse can enable virtual tourism, allowing users to explore and experience iconic landmarks, historical sites, and natural wonders from around the world. It can also simulate fictional or fantastical environments, offering unique and immersive travel experiences.

Figure 9: Tourism and the Metaverse (https://www.indianeagle.com/traveldiary/metaverse-tourism/)

9- Industry 4.0 within the Metaverse

Technology does not exist for its own sake but to make people's everyday lives better.. In a digital environment (digital twin), problems can be found, analyzed, and fixed quickly – or better yet, discovered before they arise. And all of this will happen in a world enabling a whole new level of collaboration. Where people can break the barriers of distance and work together across countries and continents as if they were together in the very same room, in front of the very same machines or objects. Where everyone can try out new ideas easily and quickly – and where innovation takes off. The Metaverse is where virtual reality supports people who are working hands on, on site.

10- Military defense and Metaverse

The environment in which the militaries are operating is changing at pace. Rapidly proliferating, increasingly multipolar threats means the battlespace of tomorrow will be more contested and congested than ever before. The military metaverse would be a digital environment that enables military personnel to train for future. It would be a virtual environment where soldiers (and algorithms) can interact with each other and simulated enemies. The metaverse would be based on data collected from real-world environments, and it would allow soldiers to experience a wide range of combat scenarios. It would also be a valuable tool for commanders, to plan, simulate and execute complex operations virtually and then use these digital playbooks for automated responses.

The Metaverse and Immersive Technologies

Concepts, Approaches, Challenges and Socio-Economical Impacts

Rachid Gherbi, Computer Science Department, Université Paris Saclay, France

Figure 10: Robot digital twin and Metaverse (https://innovateenergynow.com/resources/how-far-can-the-industrial-metaverse-go)

These are just a few examples, and the potential applications of the Metaverse extend beyond what is listed here. The true power of the Metaverse lies in its ability to create new opportunities for immersion, interaction, collaboration, and exploration across various domains.

Figure 11: Military training with Metaverse (https://dras.in/metaverse-for-armed-forces/)

The Metaverse and Immersive Technologies

Concepts, Approaches, Challenges and Socio-Economical Impacts

Rachid Gherbi, Computer Science Department, Université Paris Saclay, France

Is the Metaverse completely safe and benefical?

While the Metaverse holds great potential, there are also potential dangers and risks associated with its development and use. Among the problems cited in the previous section, here are some other concerns related to the Metaverse:

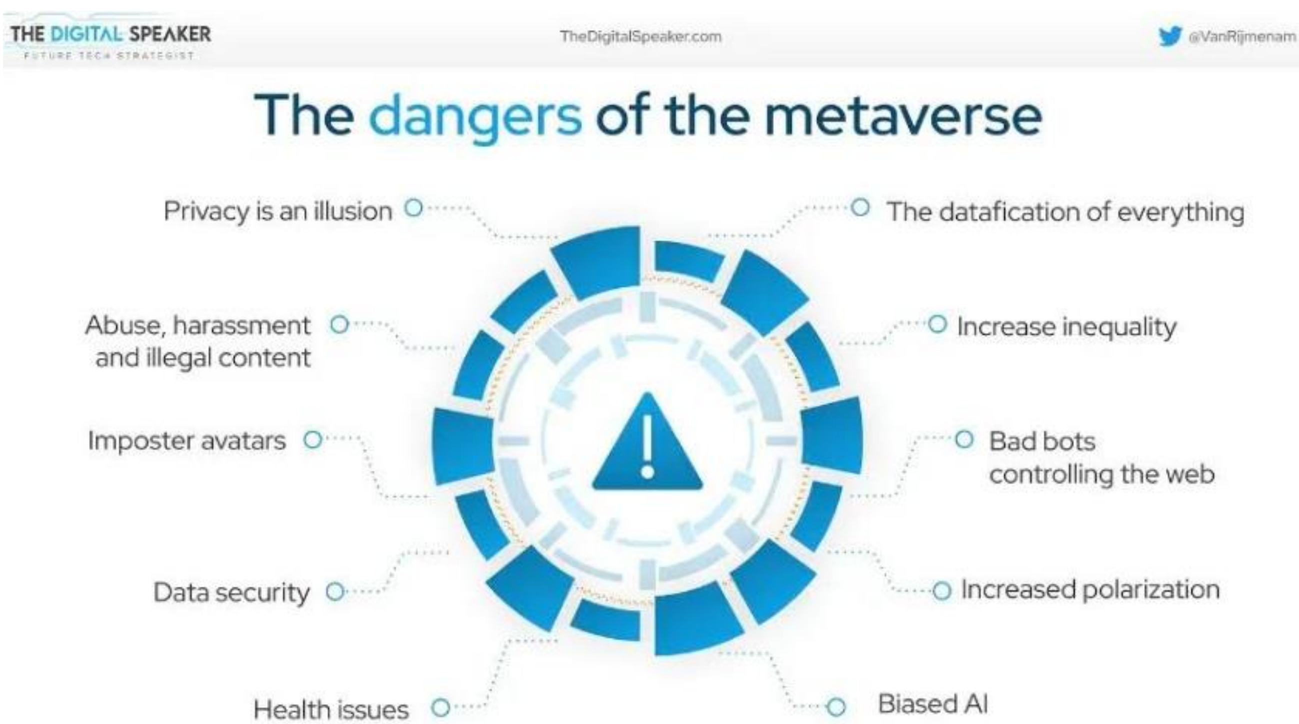


Figure 12: the danfers of Metaverse (https://www.thedigitalspeaker.com/what-are-the-dangers-of-the-metaverse/)

The Metaverse and Immersive Technologies

Concepts, Approaches, Challenges and Socio-Economical Impacts

Rachid Gherbi, Computer Science Department, Université Paris Saclay, France

1- Cybersecurity Threats: The interconnected nature of the Metaverse makes it susceptible to cyberattacks. Malicious actors could exploit vulnerabilities in the system, leading to unauthorized access, data manipulation, or disruption of services. Protecting the Metaverse's infrastructure and user data from cyber threats is crucial.

2- Digital Addiction and Behavioral Issues: The immersive and engaging nature of the Metaverse may lead to excessive use and addiction. Spending significant amounts of time in virtual environments could have negative impacts on mental health, social interactions, and physical well-being if not properly managed.

3- Virtual Identity Manipulation and Fraud: In the Metaverse, users create digital representations of themselves, known as avatars. This opens the possibility of identity manipulation, impersonation, and fraudulent activities. Ensuring the authenticity and trustworthiness of digital identities is essential to prevent misuse.

4- Economic Inequality and Exploitation: The Metaverse may exacerbate existing economic disparities if access and participation are limited to those who can afford the necessary technology and resources. There is a risk of wealth concentration, digital exploitation, and unequal distribution of opportunities and benefits within the Metaverse.

5- Legal and Jurisdictional Issues: The global nature of the Metaverse raises questions about legal frameworks, jurisdiction, and governance. Determining who has jurisdiction over virtual spaces, resolving disputes, protecting intellectual property rights, and enforcing laws can be complex in a decentralized and borderless environment.

These dangers highlight the importance of responsible development, robust security measures, clear policies, and active user education and engagement to mitigate risks and create a safe and inclusive Metaverse experience.

Visual Computing Magazine

The Metaverse and Immersive Technologies

Concepts, Approaches, Challenges and Socio-Economical Impacts

Rachid Gherbi, Computer Science Department, Université Paris Saclay, France

Companies that build the Metaverse

Several companies are actively working on Metaverse-related technologies and initiatives. Here are some notable companies in the field:

1- Facebook (Meta): Facebook, now known as Meta, has made significant investments in the development of the Metaverse. They have announced their vision for the Metaverse and are working on various projects, including the development of virtual reality (VR) technologies like Oculus, social VR experiences, and virtual collaboration tools.

2- Apple: This is the era where digital realms are becoming as significant as the physical, and in that regard, the Metaverse stands out as the next frontier in virtual exploration and interaction. Now, there's Apple's Vision Pro VRAR Headset, another groundbreaking innovation poised to redefine digital experiences. The latest in Apple technology is not just a new gadget. It represents Apple's ambitious entry into the Metaverse as it aims to merge reality with digital vastness seamlessly.

3- Epic Games is the creator of the popular game engine, Unreal Engine, and the widely successful game, Fortnite. They have shown interest in the Metaverse concept and are actively working on technologies and partnerships to bring the Metaverse to life. They have also announced the creation of the "Epic Universe," a virtual platform that aims to connect different interactive experiences.

4- Roblox is a user-generated content platform that allows users to create and play games within its virtual world. It has gained significant popularity and has been described as a precursor to the Metaverse due to its focus on user creativity and social interaction.

5- Microsoft has been investing in virtual reality, augmented reality, and mixed reality technologies with products like the HoloLens. They are exploring the potential of the Metaverse and have expressed their interest in developing tools and platforms that enable Metaverse experiences.

6- Unity Technologies is another popular game engine and development platform used by many companies to create games and interactive experiences. They are actively involved in the development of Metaverse technologies and tools, and their platform is often used to build Metaverse-related applications.

7- NVIDIA is a leading provider of graphics processing units (GPUs) and AI technologies. They are involved in the development of technologies that power virtual reality, augmented reality, and real-time graphics, which are crucial components of the Metaverse.

8- Decentraland is a blockchain-based virtual world where users can create, buy, and sell virtual assets and experiences. It is built on Ethereum and operates on a decentralized model, allowing users to have ownership and control over their digital assets.

It's important to note that the Metaverse space is evolving rapidly, and new companies and initiatives may have emerged. Additionally, large tech companies, startups, and even non-tech companies are increasingly exploring the potential of the Metaverse, so the landscape is likely to continue expanding in the coming years.

International Universities and Schools dealing with Metaverse and Immersives Technologies

Universities and schools also play a significant role in exploring and leveraging the potential of the Metaverse. While I cannot provide an exhaustive list of all the universities and schools involved, here are a few examples of academic institutions that have shown interest or taken initiatives related to the Metaverse :

1. Stanford University has been involved in research and development related to virtual reality (VR) and augmented reality (AR) technologies. Their Virtual Human Interaction Lab (VHIL) explores the social and psychological impacts of VR and AR and how they can be applied to various fields, including education and communication.

2. MIT Media Lab focuses on innovative research and interdisciplinary collaboration in various areas of media and technology. They have conducted research on virtual and augmented reality, interactive storytelling, and the future of media, which are relevant to the Metaverse.

3. Harvard University has explored the use of virtual reality in education and has conducted research on VR's potential to enhance learning experiences. They have also been involved in projects related to virtual museums and cultural heritage preservation.

4. University of Southern California (USC) has a strong presence in the field of interactive media and has been involved in research on VR, AR, and mixed reality (MR) technologies. They have developed projects related to virtual storytelling, immersive education, and virtual environments for cultural experiences.

5. Carnegie Mellon University has a renowned Entertainment Technology Center (ETC) that focuses on the intersection of technology, art, and entertainment. They have conducted research on virtual reality, game design, and interactive experiences, which are relevant to the Metaverse.

6. NYU Tandon School of Engineering has been involved in research on virtual reality and augmented reality technologies. They have explored applications of VR/AR in fields such as education, healthcare, and urban planning.

7. Asian universities
8. Arabic and African universities
9. European universities

These are just a few examples, and many other universities and schools around the world are likely exploring the potential of the Metaverse in education, research, and other domains. The involvement and initiatives can vary based on individual faculty, departments, research centers, and the overall focus of the institution.

The Metaverse and Immersive Technologies

Concepts, Approaches, Challenges and Socio-Economical Impacts

Rachid Gherbi, Computer Science Department, Université Paris Saclay, France

The Metaverse in Arabic and African Regions

Regarding the Metaverse and its specific application in these regions, as of my present knowledge, there is no specific information available on the development or implementation of the Metaverse as a whole technology. However, it is worth noting that the adoption and development of the Metaverse can vary from country to country based on factors such as technological infrastructure, economic conditions, and government initiatives.

Many countries from these regions, may face challenges in terms of infrastructure development, internet connectivity, and access to advanced technologies. These factors can impact the adoption and realization of the Metaverse concept in the country. However, it is important to note that technological advancements can change rapidly, and there may have been developments.

The specific developments and initiatives related to the Metaverse in Arabic and African countries may vary. However, it is worth noting that the concept of the Metaverse has gained global attention, and various countries, have shown interest in exploring its potential. These countries have been actively investing in digital transformation and emerging technologies. Some countries, such as the United Arab Emirates (UAE), have made significant efforts to foster innovation and technological advancements. For example, the UAE launched the "UAE Strategy for the Fourth Industrial Revolution" to promote the adoption of emerging technologies like artificial intelligence (AI), blockchain, and augmented reality (AR)/virtual reality (VR). Dubai, in particular, has positioned itself as a hub for innovation and futuristic technologies. The Dubai Future Foundation, through initiatives like the Dubai Future Accelerators, has been supporting technology-driven projects and fostering collaborations with startups and tech companies.

The Metaverse and Immersive Technologies

Concepts, Approaches, Challenges and Socio-Economical Impacts

Rachid Gherbi, Computer Science Department, Université Paris Saclay, France

Arabic and African countries also have a growing gaming and technology industry, and there is a potential for the Metaverse to gain traction in these regions. Additionally, these countries have shown interest in utilizing virtual reality (VR) and augmented reality (AR) technologies in areas such as tourism, education, and cultural preservation. These technologies are closely related to the Metaverse concept.

The specific developments and initiatives related to the Metaverse in Africa may vary. However, Africa, like other regions, has shown interest in the potential of emerging technologies and digital transformation. For example, countries like Kenya, Nigeria, South Africa, and Rwanda have witnessed significant growth in their technology sectors and entrepreneurial ecosystems.

While the Metaverse is still an evolving concept globally, the immersive and interactive nature of the Metaverse can provide unique opportunities to bridge gaps, connect people, and promote digital inclusion across the continent.
In recent years, there has been an increase in African tech startups and innovators working on VR, AR, and other related technologies. These companies are contributing to the exploration and development of immersive experiences and virtual environments that could align with the Metaverse concept.

The Metaverse vs Google & Amazon

Amazon and Google are major technology companies that have shown interest and made investments in various areas related to the Metaverse. Here are some relevant initiatives and technologies they have been involved with as of my knowledge:

1. Amazon Web Services (AWS): AWS provides cloud computing services and has offerings related to VR and AR, and 3D content rendering. AWS has been used by companies and developers to build and deploy Metaverse-related applications and services.

2. Amazon Sumerian: Amazon Sumerian is a platform for creating and publishing AR, VR, and 3D experiences. It offers tools for building immersive applications and supports various devices and platforms.

3. Google VR and AR: Google has invested in virtual reality and augmented reality technologies. They developed platforms like Google Cardboard and Daydream for VR experiences, as well as ARCore for building augmented reality applications on Android devices.

4. Google Poly: Google Poly is a platform for discovering, previewing, and downloading 3D models and assets. It provides a library of 3D content that can be used in virtual worlds and AR applications.

5. Google Earth VR: Google Earth VR allows users to explore and navigate a virtual representation of the Earth using VR technology. It provides an immersive experience of visiting different locations and landmarks.

It's worth noting that these companies involvement in the Metaverse space can evolve and expand over time, and they may have launched new initiatives or projects.

Future Opportunities of Metaverse and Immersive Technologies

Future of the Metaverse holds numerous opportunities across various domains. Here are some potential future opportunities associated with the Metaverse:

1. Social Interaction and Communication: The Metaverse can enhance social interactions by providing immersive and engaging virtual environments. People can connect, collaborate, and communicate in new and exciting ways, fostering global connections and breaking down geographical barriers.

2. Work and Business: The Metaverse has the potential to transform the way we work and conduct business. Virtual offices, meetings, and collaboration spaces can enable remote work and global teamwork. Virtual marketplaces and economies within the Metaverse can create new opportunities for entrepreneurship and commerce.

3. Education and Training: The Metaverse can revolutionize education and training by offering immersive and interactive learning experiences. Virtual classrooms, simulations, and personalized learning environments can enhance engagement, accessibility, and effectiveness of educational programs.

4. Entertainment and Gaming: The entertainment and gaming industries can greatly benefit from the Metaverse. It can provide immersive gaming experiences, virtual concerts, live events, and interactive storytelling, blurring the line between real and virtual worlds.

5. Healthcare and Therapy: The Metaverse can be leveraged for telemedicine, remote patient monitoring, and therapy services. Virtual environments can facilitate medical training, simulations, and patient education, enabling innovative approaches to healthcare delivery.

6. Cultural Preservation and Tourism: The Metaverse can help preserve cultural heritage by creating virtual replicas of historical sites and artifacts. It can offer virtual tourism experiences, allowing people to explore iconic locations and experience different cultures from anywhere in the world.

7. Art, Creativity, and Expression: The Metaverse can be a canvas for artists and creators to exhibit their work and engage with audiences in immersive ways. It can enable collaborative art projects, virtual galleries, and interactive installations that redefine the boundaries of artistic expression.

8. Personal Development and Wellness: The Metaverse can provide opportunities for personal growth, wellness, and self-improvement. Virtual environments can offer meditation, mindfulness, and fitness experiences, promoting mental and physical well-being.

9. Environmental Sustainability: The Metaverse can contribute to environmental sustainability by reducing the need for physical travel and infrastructure. Virtual experiences can minimize carbon footprints and support eco-friendly practices.

These opportunities are just a glimpse of the potential the Metaverse holds. As technology advances and the concept evolves, new and unforeseen opportunities may arise, paving the way for exciting innovations and advancements in various aspects of our lives.

The Dark Side of the Metaverse

While the Metaverse holds great potential, it's important to consider potential challenges and risks associated with its development and use. Here are some concerns often discussed as the potential dark side of the Metaverse:

1. Privacy and Security Risks: The Metaverse collects vast amounts of personal data and user interactions, raising concerns about privacy and security. Unauthorized access, data breaches, identity theft, and surveillance are potential risks that need to be addressed to protect user information.

2. Addiction and Escapism: The immersive and engaging nature of the Metaverse can lead to excessive use and addiction, potentially resulting in negative impacts on mental health, physical well-being, and real-world relationships. Individuals may become overly reliant on virtual experiences, leading to a disconnection from the physical world.

3. Exploitation and Fraud: In virtual environments, there is a risk of virtual identity manipulation, fraud, and scams. Impersonation, theft of virtual assets, and fraudulent transactions can occur, requiring robust measures to ensure authenticity, trust, and security within the Metaverse.

4. Inequality and Exclusion: The Metaverse could exacerbate existing inequalities, creating a digital divide between those who can afford the necessary technology and access and those who cannot. Economic disparities may result in unequal participation, opportunities, and benefits within the Metaverse.

5. Ethical Dilemmas and Content Regulation: The Metaverse raises ethical questions regarding content creation, moderation, and regulation. Addressing hate speech, harmful or inappropriate content, and ensuring inclusivity and diversity can be complex in a decentralized and vast virtual environment.

6. Social Isolation and Disconnection: Spending excessive time in virtual environments may lead to social isolation, disconnection from the physical world, and a decline in real-world relationships and community bonds. Over-reliance on virtual interactions could have negative social implications.

7. Loss of Authenticity and Manipulation: In the Metaverse, digital representations (avatars) can be manipulated or artificially created. This raises concerns about authenticity, trust, and the potential for deception, as distinguishing between real and virtual identities may become more challenging.

8. Environmental Impact: The energy consumption and carbon footprint associated with running the infrastructure required for the Metaverse could have environmental implications. The data centers, servers, and high-performance computing needed to support the Metaverse can contribute to increased energy consumption and greenhouse gas emissions.

These concerns highlight the need for responsible development, comprehensive regulations, user education, and ongoing discussions to address and mitigate the potential negative impacts of the Metaverse.

The Metaverse and Immersive Technologies

Concepts, Approaches, Challenges and Socio-Economical Impacts

Rachid Gherbi, Computer Science Department, Université Paris Saclay, France

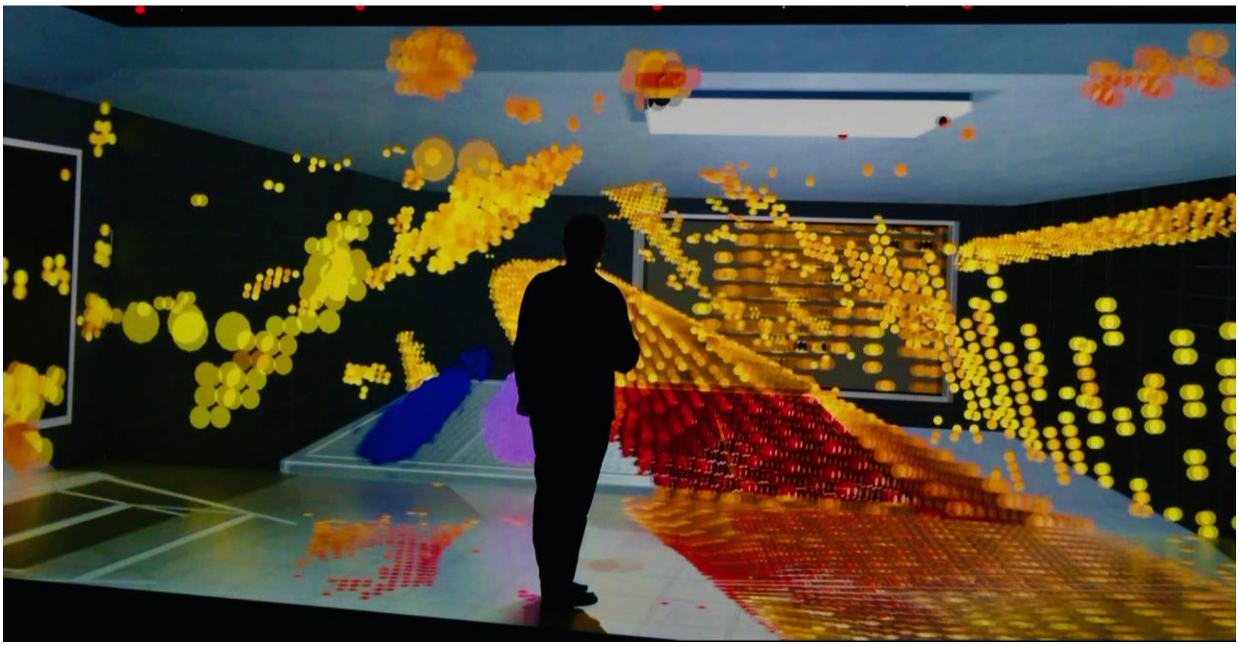
Bibliography

1. John David N. Dionisio, William G. Burns III, & @Richard Gilbert (2013). 3D virtual worlds and the metaverse: Current status and future possibilities. ACM Computing Surveys (CSUR), 45(3), 1-38.
2. Sang-Min Park & Young-Gab Kim (2022). A metaverse: Taxonomy, components, applications, and open challenges. IEEE access, 10, 4209-4251.
3. Yogesh K Dwivedi, Dr Laurie Hughes, and all, (2022). Metaverse beyond the hype: Multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy. International Journal of Information Management, 66, 102542.
4. Haihan Duan, Jiaye Li, Sizheng Fan, Zhonghao Lin, Xiao Wu & Wei Cai (2021, October). Metaverse for social good: A university campus prototype. In Proceedings of the 29th ACM international conference on multimedia (pp. 153-161).
5. Bokyung Kye, Nara Han, Eunji Kim, Yeonjeong Park, & Soyoung Jo (2021). Educational applications of metaverse: possibilities and limitations. Journal of educational evaluation for health professions, 18.
6. Hee-soo Choi, & Sang-heon Kim (2017). A content service deployment plan for metaverse museum exhibitions—Centering on the combination of beacons and HMDs. International Journal of Information Management, 37(1), 1519-1527.
7. Jooyoung Kim (2021). Advertising in the metaverse: Research agenda. Journal of Interactive Advertising, 21(3), 141-144.
8. Gwo-Jen Hwang, & Shu-Yun Chien (2022). Definition, roles, and potential research issues of the metaverse in education: An artificial intelligence perspective. Computers and Education: Artificial Intelligence, 3, 100082.
9. Sascha Kraus, Dominik K. Kanbach, Dr. Peter Krysta, Dr. Maurice Steinhoff, & Nino Tomini (2022). Facebook and the creation of the metaverse: radical business model innovation or incremental transformation? International Journal of Entrepreneurial Behavior & Research, 28(9), 52-77.
10. Michael Bourlakis, Savvas Papagiannidis & Feng Li (2009). Retail spatial evolution: paving the way from traditional to metaverse retailing. Electronic commerce research, 9, 135-148.
11. Ethical concerns in contemporary virtual reality and frameworks for pursuing responsible use. Urooj Raja & Reem Al-Baghli - 2025 - Frontiers in Virtual Reality 6.
12. The Metaverse: Surveillant Physics, Virtual Realist Governance, and the Missing Commons. Andrew McStay - 2023 - Philosophy and Technology 36 (1):1-26.
13. If the Metaverse Becomes an Ontological Event.Tingyang Zhao - 2022 - Journal of Human Cognition 6 (1):3-17.

Visual Computing Magazine

Immersive Volumetric Point Cloud Manipulation for Cultural Heritage

Rafik Drissi, DIGI ROOTS XR, Ronan Gaugne, Univ Rennes, Inria, CNRS, IRISA, France, Théophane Nicolas, Inrap, UMR 8215 Trajectoires, France, Valérie Gouranton, Univ Rennes, INSA Rennes, Inria, CNRS, IRISA, France



Abstract

In this article, we present a framework for an immersive and interactive 3D manipulation of volumetric point clouds in virtual reality. The framework was designed to meet the needs of cultural heritage experts such as archaeologists or curators for use on cultural heritage artifacts. We propose a display infrastructure associated with a set of tools that allows users from the cultural heritage domain to interact directly with the point clouds within their study process. The resulting framework allows an immersive navigation, interaction and real time segmentation.

Index Terms— Virtual Reality; Point Cloud; Volume Rendering; Cultural Heritage;

I. INTRODUCTION

Cultural heritage professionals use a wide variety of digital data in conducting their studies. These come from different sources and can be used for the visualization and manipulation of 3d surface or volume representations. Computed tomography scanning is an important source of volume data, and in a Cultural Heritage context, is used as a nondestructive means of studying archaeological artefacts [9], [18], [17].

Immersive Volumetric Point Cloud Manipulation for Cultural Heritage

Rafik Drissi, DIGI ROOTS XR, Ronan Gaugne, Univ Rennes, Inria, CNRS, IRISA, France, Théophane Nicolas, Inrap, UMR 8215 Trajectoires, France, Valérie Gouranton, Univ Rennes, INSA Rennes, Inria, CNRS, IRISA, France

It becomes possible to assess the degradation of the artefacts, and serves as a guide in the restoration process for the implementation of a suitable protocol and the deployment of preservation means before any destructive micro-excavation operation. For such cases, some Cultural Heritage specialists have access to two dimensional, non-immersive standard tools derived from the uses in the medical field for harnessing the potential of their medical imaging data, however these tools are not designed for ease of use and oftentimes require an amount of skill that cultural heritage specialists lack, particularly when it comes to segmentation operations. The discrete nature of the three-dimensional grid structure of DICOM (Digital imaging file format) data, has oriented our research towards the use of point clouds as the main representation of the volume data, it is a representation that is supported by numerous, readily available third-party software. In this paper, we design and implement a framework that provides a dedicated solution to cultural heritage professionals to exploit their CT-scan data by themselves, through intuitive and easy to use real time interactive and immersive 3d visualization and segmentation tools and illustrate these features with three sample use cases.

II. Related Works

There have been numerous works attempting to harvest the potential of the 3d medical data in ct-scan DICOM files since the introduction of the x-ray CT scanner in the 1970's, in [7], Herlan and Liu presented a three-dimensional voxel-based visualization of Human Organs from computed tomography. The quality of the images has steadily improved enabling the role of the CT scanner to expand into the field of cultural heritage. Indeed, this technology provides internal viewing of the material that helps to identify the internal content and to optimize the preparation of the excavation, as pointed out in the works [18], and [15]. But this kind of image use remains still in a limited proportion due to the difficulties of exploiting the obtained digital data.

Visual Computing Magazine

Immersive Volumetric Point Cloud Manipulation for Cultural Heritage

Rafik Drissi, DIGI ROOTS XR, Ronan Gaugne, Univ Rennes, Inria, CNRS, IRISA, France, Théophane Nicolas, Inrap, UMR 8215 Trajectoires, France, Valérie Gouranton, Univ Rennes, INSA Rennes, Inria, CNRS, IRISA, France

Also, the use of point cloud data has been researched and used in a variety of fields, in [3] Chougule and al explored three-dimensional point cloud generations from CT scan images for bio-cad modeling, the resulting point cloud data was plotted with the help of Imaging software thus presenting very basic visualization and no interaction features. Hanel and Al explored the benefits of IVEs (Immersive Virtual Environments) for the visual analysis of spatial data in [8], but concluded that the full potential of IVEs cannot be utilized due to a strong necessity of interactivity in such environments and the lack in computer performance, our work uses volumetric point clouds to achieve both visualization and interaction goals.

Many projects relying on point clouds for the encoding of data uses output from LIDAR scanning or photogrammetry techniques, e.g. [16], [11] in a geological context, and [12] in a cultural heritage context, thus only getting interested in data of surface representations and not that of volumetric data as explored in this project. Such projects can be found in the 2014 IEEE 3DUI contest in the theme of 3D annotation task of point-could data sets, in [2], Cabral and al implemented a desktop system allowing for a bi-manual gesture interface to navigate through the 3D point cloud data within a 3D modelling tool to explore, select and/or annotate points, As well as Bacim and al who proposed A free-hand gesture user interface for 3D point cloud annotation, in [1]. In [19], Veit and Capobianco presented a framework for selection and annotation of point clouds through a tracked multi-touch device that combines 2D and 3D interaction techniques in a single device, the selection was made through an Algorithm for progressive refinement of point clouds into a tree of sub-point clouds to which individual points could be added or removed.

Among the works providing immersive selection of points within a point cloud, Krammes et al., in [10], proposed an implementation of a selection of surface point clouds through a mobile device. In [6], Gaugne and al. presented a framework for an immersive and interactive 3D manipulation of large point clouds, which was limited to surface representations and to cross-sections segmentation options.

Visual Computing Magazine

Immersive Volumetric Point Cloud Manipulation for Cultural Heritage

Rafik Drissi, DIGI ROOTS XR, Ronan Gaugne, Univ Rennes, Inria, CNRS, IRISA, France, Théophane Nicolas, Inrap, UMR 8215 Trajectoires, France, Valérie Gouranton, Univ Rennes, INSA Rennes, Inria, CNRS, IRISA, France

Similarly to [10] and [1], Montano-Murillo et al explored in [14], a Hybrid 3D/2D Multi-target Selection Technique for Dense Virtual Environments, thus working on volumetric point clouds and focusing on providing a hybrid workflow combining mid-air and tablet interactions, but only focused on advantages of the hybrid selection technique. In [13], Maloca and al. presented a Virtual Reality Volume Rendering of Original Optical Coherence Tomography Point-Cloud Data Enhanced with Real-Time Ray Casting, that allowed the user to import DICOM files into the VR program, interact with a light source and operate cross sections on the 3d representation but did not allow for segmentation of the 3d data or simultaneous manipulation of multiple DICOM files.

In summary, the previous works are lacking in the integration of the medical imaging files as the input data, in the simultaneous and independent manipulation of point clouds from multiple sources, in the focus on Cultural heritage use and in the real time segmentation and extraction of point clouds.

III. FRAMEWORK DESIGN

Main features :

Our collaboration with the archaeologists and curators involved in this project has resulted in the selection of a number of features that the user must be able to conduct, which are the following: The user can view cross-sections of multiple point clouds (Figure 1, thus revealing a view on the interior in any angle or position. He can isolate the point clouds, select and extract any subset of points from the point clouds, and perform all operations on the newly generated point cloud. The user can change the point clouds color through a handy color picker, thus modifying the perception of the point clouds (Figure 2), this feature allows to perceive surface details (topography) and identify composition of the point clouds. It can even allow to identify different wood patterns or engravings (Figure 3).

Immersive Volumetric Point Cloud Manipulation for Cultural Heritage

Rafik Drissi, DIGI ROOTS XR, Ronan Gaugne, Univ Rennes, Inria, CNRS, IRISA, France, Théophane Nicolas, Inrap, UMR 8215 Trajectoires, France, Valérie Gouranton, Univ Rennes, INSA Rennes, Inria, CNRS, IRISA, France

The user can vary the viewpoints and investigate the 3-dimensional space, 1:1 scale for a realistic analysis, miniature for an overview and magnified to analyze morphological details, holes, empty spaces and hidden parts within the object (Figure 4). The user can integrate and simultaneously study multiple point clouds (especially in heterogeneous material, (Figure5), bloc excavation (of an incineration for example) provides a rare opportunity for a direct comparison between these elements (usually stored in distant places or studied at different times).

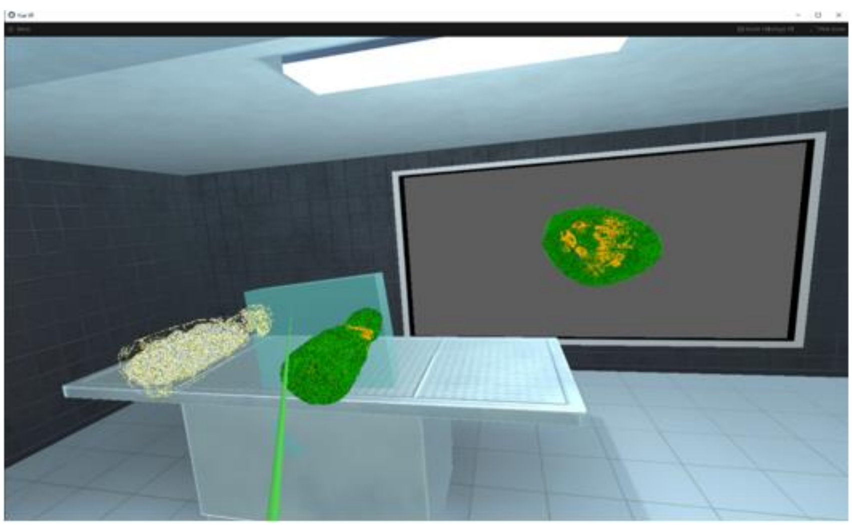


Figure 1: Interactive cross-section: the cutting plan in blue is positioned interactively on an object and the resulting cross section is displayed on the back wall

The user's interactions are optimized in favor of enhanced intuitiveness and ease of use by non-computer scientists. It is possible to grab and manipulate point clouds ergonomically, to handle an object in any direction, internally and at any scale, thus providing relevant support for scientific study.

Visual Computing Magazine

Immersive Volumetric Point Cloud Manipulation for Cultural Heritage

Rafik Drissi, DIGI ROOTS XR, Ronan Gaugne, Univ Rennes, Inria, CNRS, IRISA, France, Théophane Nicolas, Inrap, UMR 8215 Trajectoires, France, Valérie Gouranton, Univ Rennes, INSA Rennes, Inria, CNRS, IRISA, France

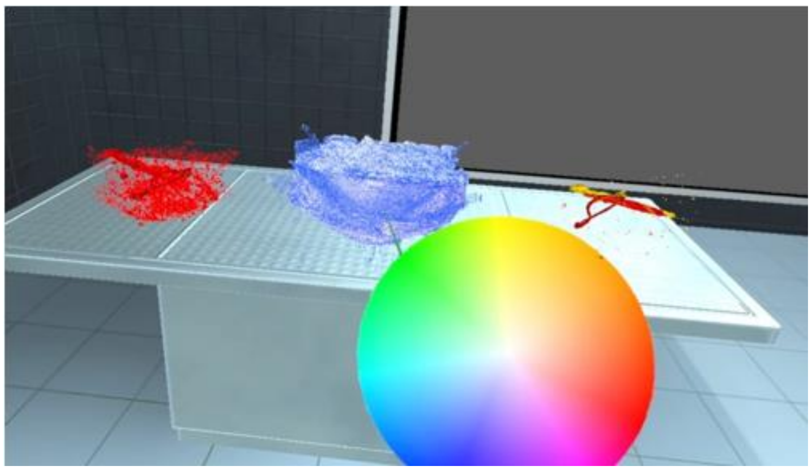


Figure 2: Color selection

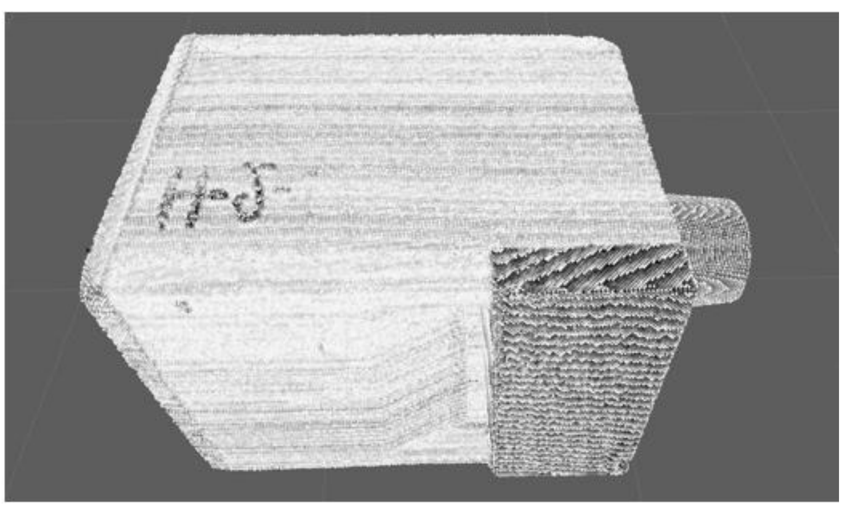


Figure 3: Point cloud of wooden whistle module with apparent wood grain and inscription.

Immersive Volumetric Point Cloud Manipulation for Cultural Heritage

Rafik Drissi, DIGI ROOTS XR, Ronan Gaugne, Univ Rennes, Inria, CNRS, IRISA, France, Théophane Nicolas, Inrap, UMR 8215 Trajectoires, France, Valérie Gouranton, Univ Rennes, INSA Rennes, Inria, CNRS, IRISA, France

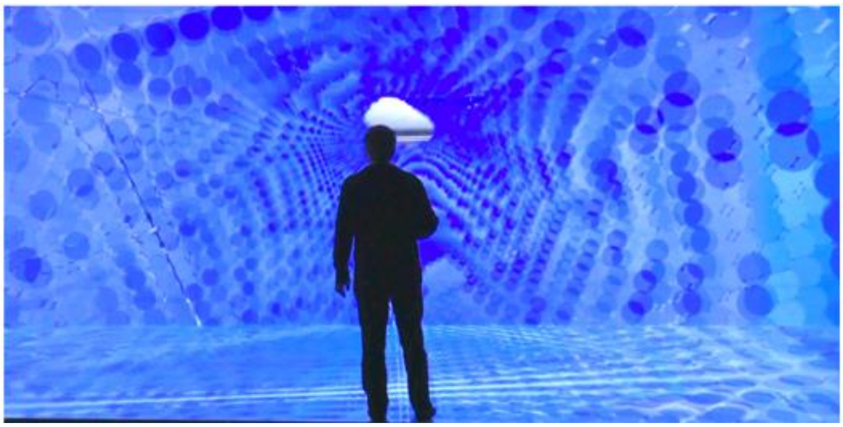


Figure 4: The user navigates inside the magnified point cloud, here in the cavity of a feline long bone.

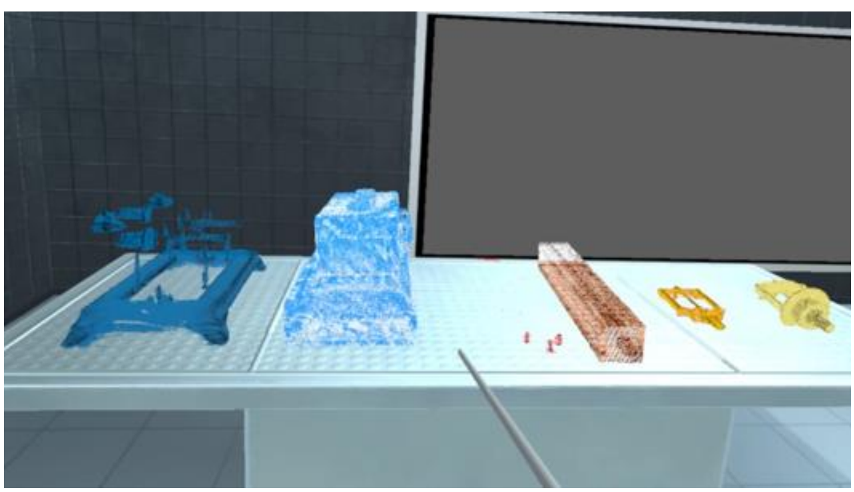


Figure 5: Visualization of the different point clouds of a complex machine.

Immersive Volumetric Point Cloud Manipulation for Cultural Heritage

Rafik Drissi, DIGI ROOTS XR, Ronan Gaugne, Univ Rennes, Inria, CNRS, IRISA, France, Théophane Nicolas, Inrap, UMR 8215 Trajectoires, France, Valérie Gouranton, Univ Rennes, INSA Rennes, Inria, CNRS, IRISA, France

IV. IMPLEMENTATION

A. DICOM to point cloud converter:

This first part of the framework converts the medical imaging DICOM files and outputs point clouds in ASCII or Binary Stanford Triangle Format (PLY extension) for readability and file size optimization. The converter combines two segmentation techniques to generate several point clouds. The first segmentation technique is based on the density. To each point cloud is associated a range of density. The second segmentation technique is based on the setting of a rectangular area in the DICOM images. Only the points within the area will be considered for the point cloud generation. Both techniques can be combined to generate accurate point clouds. The input to the converter is medical imaging DICOM files. A list of density thresholds and optional position are also fed into the software in order to do a primary segmentation in the generation of the initial point clouds. The position limits define the rectangular area in the DICOM images to restrict the points considered by the converter.

This segmentation technique is useful to dissociate elements of the same density that were scanned simultaneously, as in Figure 6 where a high-density metal piece in the left part of a cremation urn is separated from other metal pieces. It also optimizes the process by making the converter computation faster and generating smaller point clouds objects. The grey scale colors represent the variation in density of the points within the defined density range. Which makes it possible to visualize details such as oxidation, wood grain, stains, painting or engravings as shown in Figure 3.

The converter outputs point clouds into high compatibility PLY file standard, thus allowing for use in third-party software such as Cloudcompare.

Visual Computing Magazine

Immersive Volumetric Point Cloud Manipulation for Cultural Heritage

Rafik Drissi, DIGI ROOTS XR, Ronan Gaugne, Univ Rennes, Inria, CNRS, IRISA, France, Théophane Nicolas, Inrap, UMR 8215 Trajectoires, France, Valérie Gouranton, Univ Rennes, INSA Rennes, Inria, CNRS, IRISA, France

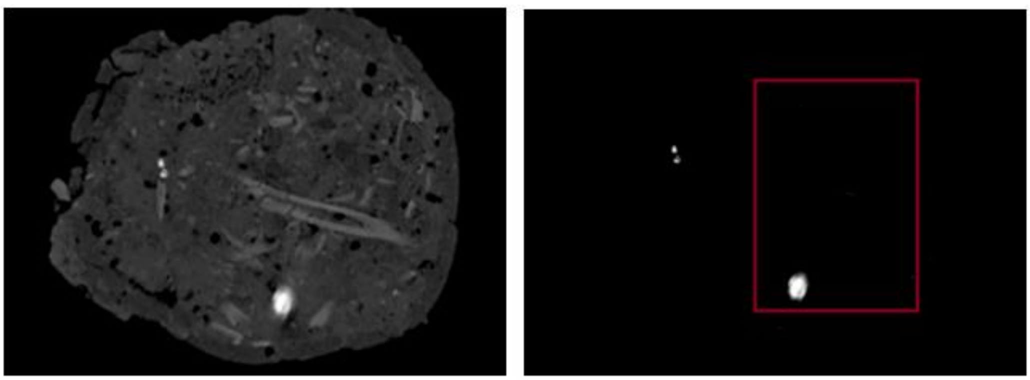


Figure 6: From left to right: Dicom visualization of original image, and Dicom image after criteria selection applied (position in red)

B. Immersive Visualization and Interaction environment:

The second part of the framework takes the generated point clouds as an input and provides the visualization and interaction features. We will now present it in three parts: Rendering, Manipulation and Segmentation.

1) Rendering: The point clouds are imported as compute buffers, i.e. storage that can be accessed both by CPU and GPU. Each point cloud compute buffer (defined by a density range and positional properties) receives a bounding box and a material using a shader, which renders the point cloud in volume splatting. The material gives the points a uniform modifiable radius, and two colors from which to form a gradient. Initially, as presented in the previous section, all imported point clouds are grey and cannot be differentiated by their density according to color alone. In order to do so, our framework allows to remap the initial grey color of each point to a color defined by a linear gradient between two color references. The difference between the colors as imported and the result of the color mapping between yellow (255, 255, 0) and red (255,0,0) is shown in Figure 7.

Immersive Volumetric Point Cloud Manipulation for Cultural Heritage

Rafik Drissi, DIGI ROOTS XR, Ronan Gaugne, Univ Rennes, Inria, CNRS, IRISA, France, Théophane Nicolas, Inrap, UMR 8215 Trajectoires, France, Valérie Gouranton, Univ Rennes, INSA Rennes, Inria, CNRS, IRISA, France

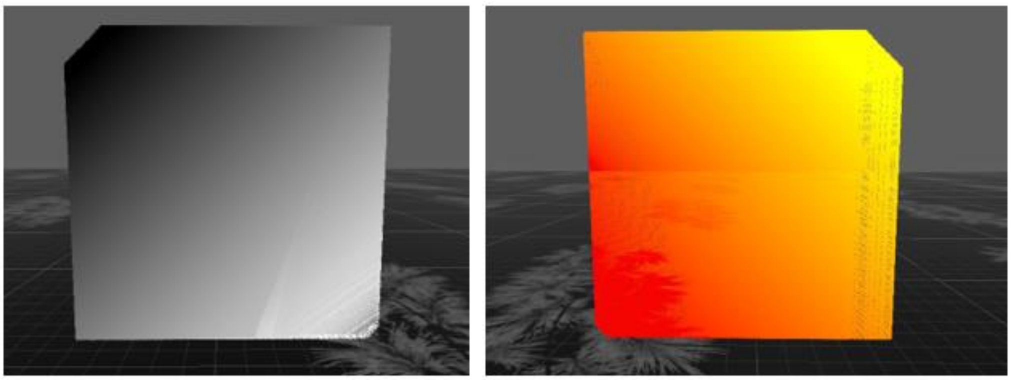


Figure 7: From left to right: Original grey point cloud and point cloud with axial gradient from yellow to red applied

This is useful as the user can easily categorize objects by color or see point clouds that are occluded within or behind another point cloud.

2) Navigation :

We designed the immersive environment with the goal of being accessible and easy to use. Thus, we use One-handed Flying navigation techniques which provides ease of use, accuracy and is easy to learn as demonstrated in [4]. It also allows for movement while manipulating an object or using the menu. The user starts in a closed environment mimicking an excavation lab as shown in Figure 8, he can at any moment toggle it off and use an unlimited and uncluttered space.

Manipulation is done using Direct HOMER (Hand-Centered Object Manipulation Extending Ray-Casting) techniques based on natural hand motions, thus providing a natural and efficient control over the selected point cloud’s position and orientation, confirmation of the selection happens by change in the selected point cloud’s color. The user can manipulate a cross-section box, which displays a cross section of the point cloud on a virtual screen on the back wall of the room.

Immersive Volumetric Point Cloud Manipulation for Cultural Heritage

Rafik Drissi, DIGI ROOTS XR, Ronan Gaugne, Univ Rennes, Inria, CNRS, IRISA, France, Théophane Nicolas, Inrap, UMR 8215 Trajectoires, France, Valérie Gouranton, Univ Rennes, INSA Rennes, Inria, CNRS, IRISA, France

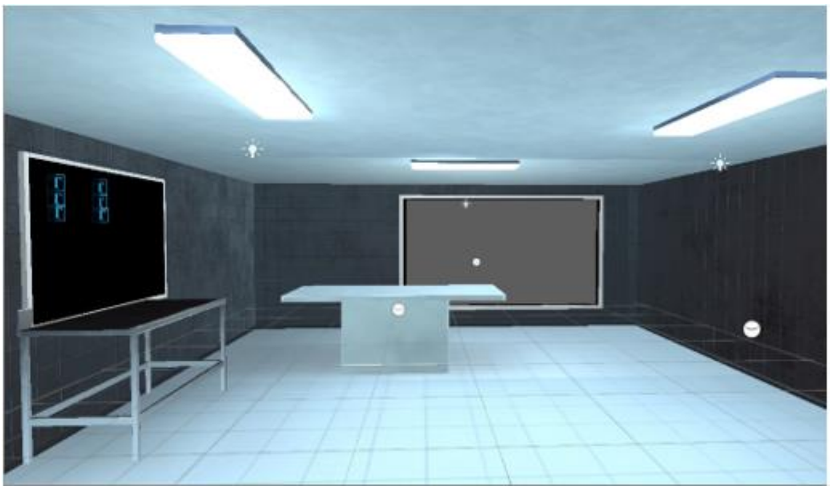


Figure 8: Empty interaction closed environment

This feature can be useful before the user starts manipulating the point clouds to see how they are disposed. The scale feature allows for making point clouds as big or as small as the user wants them to. Making the point cloud bigger allows for the user to inspect the cavities of the object. This feature also allows to change the size of the selection and segmentation sphere, thus allowing bulkier or finer selection/deselection. When the user doesn't intend to use point clouds or no longer needs selections, he/she's made, he/she can choose to delete them, thus making way for enhanced performances.

3) Segmentation: The user has the ability to make a 3d segmentation of a point cloud. This functionality allows to isolate an even smaller portion of the studied object. The user holds a selection sphere in his hand, and while the selection mode is active, all the points within the selection sphere's radius are added to or removed from a newly generated point cloud using dedicated buttons. The user can keep adding and removing points from the clone until he considers the area of interest isolated. At which point the newly generated point cloud can be interacted with as an independent point cloud and can even be reused for a new segmentation.

Immersive Volumetric Point Cloud Manipulation for Cultural Heritage

Rafik Drissi, DIGI ROOTS XR, Ronan Gaugne, Univ Rennes, Inria, CNRS, IRISA, France, Théophane Nicolas, Inrap, UMR 8215 Trajectoires, France, Valérie Gouranton, Univ Rennes, INSA Rennes, Inria, CNRS, IRISA, France

C. Technical environment:

The converter from DICOM medical imaging files to point clouds was implemented using the Python programming language, with the libraires: tkinter for the visual interface and VTK for the manipulation of the DICOM data.

As for the VR tool we use Unity version 2018.3 and rely on the point cloud importer package "Pcx" as a base. Tests are conducted on a VR Ready Desktop with the following characteristics: i7-9700k processor, 32GB DDR4-2400 Ram, Nvidia GTX 1080 GPU, and an Oculus Rift CV1 with touch controllers.

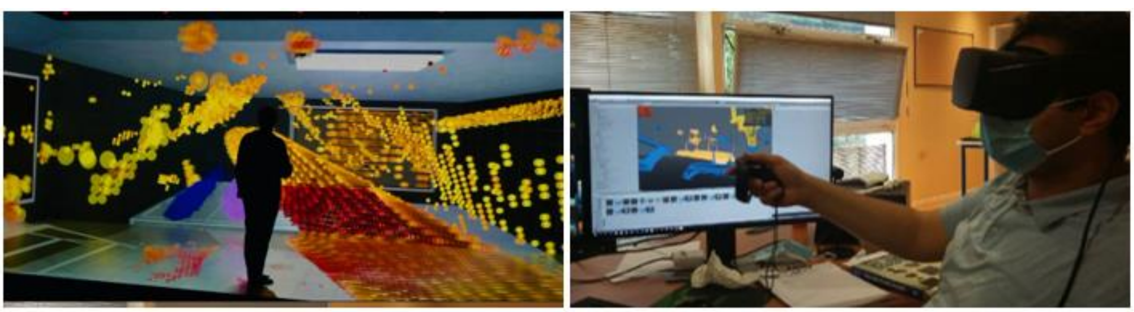


Figure 9: Use of the VR interface in the CAVE and with the HMD implementation

This tool relies on the VR plugin for Unity MiddleVR3 as a middle-ware to manage the multiple VR visualization hardware. It has been tested on common Hardware from Oculus Rift and HTC Vive head mounted displays, as well as the large CAVE-like facility presented in Figure 9 constituted of 4 screens, one floor of 10m x 3m, two lateral screens of 3m x 3m, and a main vertical screen of 10m x 3m, with more than 45M pixels displayed by 14 WQXGA laser projectors, driven by 7 PC, with the following characteristics: 2 Intel Xeon E5-2623v4 2.6 2133 4C Processors, 32GB DDR4 2400 Ram, 2 NVIDIA Quadro P6000 24GB GPU.

Immersive Volumetric Point Cloud Manipulation for Cultural Heritage

Rafik Drissi, DIGI ROOTS XR, Ronan Gaugne, Univ Rennes, Inria, CNRS, IRISA, France, Théophane Nicolas, Inrap, UMR 8215 Trajectoires, France, Valérie Gouranton, Univ Rennes, INSA Rennes, Inria, CNRS, IRISA, France

V. EXPERIMENTAL RESULT

A. Cremation urn case :

The cremation urn was pre-segmented into three density thresholds:
The fibula and the blade: density threshold from 3500 to 10 000
The bones and metal corrosion layer: density threshold from 1800 to 2500
the ceramic and sediments: density threshold from 500 to 1500

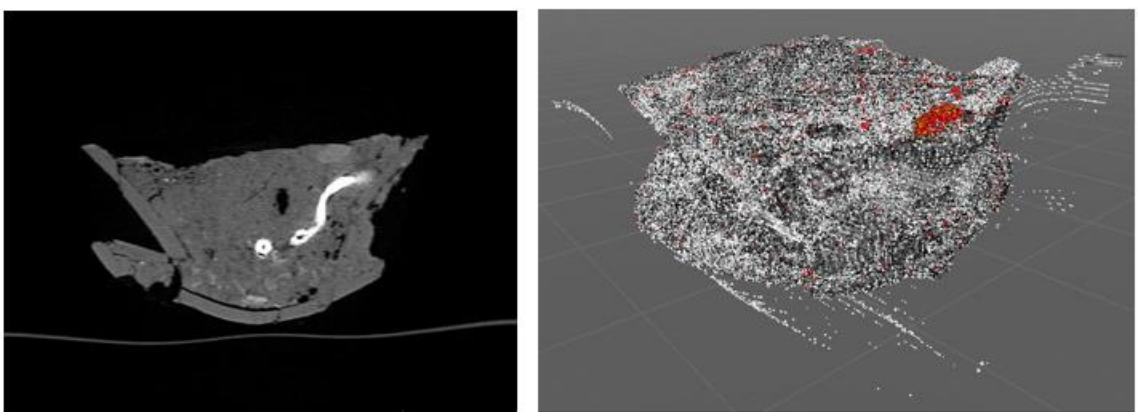


Figure 10: From left to right : Scan of cremation urn and VR visualization in the framework

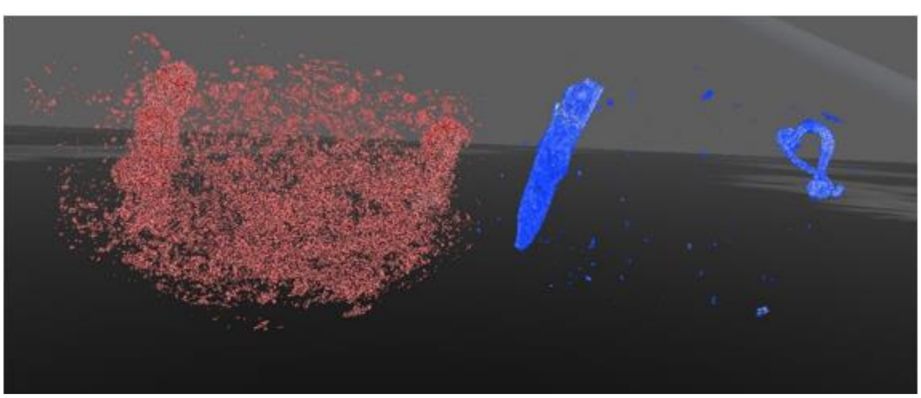


Figure 11: point clouds of metal and sediment content of the cremation urn

Visual Computing Magazine

Immersive Volumetric Point Cloud Manipulation for Cultural Heritage

Rafik Drissi, DIGI ROOTS XR, Ronan Gaugne, Univ Rennes, Inria, CNRS, IRISA, France, Théophane Nicolas, Inrap, UMR 8215 Trajectoires, France, Valérie Gouranton, Univ Rennes, INSA Rennes, Inria, CNRS, IRISA, France

The three-point clouds generated from this first step were respectively constituted of 42,483; 174,086 and 7,881,496 points. The framework allows to easily visualize and separate a bloc excavation containing multiple materials (ceramic, sediment, bone, and iron oxide) that can have relatively similar densities. We were thus able to independently manipulate the point cloud of the metallic objects as shown in Figure 11. From this last point cloud, it was straightforward to perform the segmentation of the point cloud in the virtual reality environment to extract the fibula as shown in Figure 12.

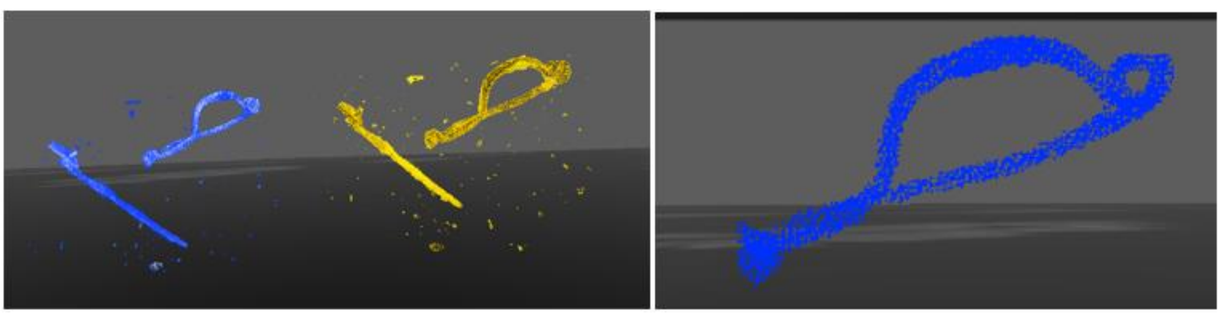


Figure 12: Left top to right: point cloud of metallic content of the cremation urn with clone, and extracted fibula point cloud

B. Mummy cat case :

The second material of the cat mummy also received a pre-segmentation into three density thresholds:
The bones: density threshold from 1600 to 10 000
The dense wrappings: density threshold from 50 to 200
The outer wrappings: density threshold from -50 to 10

The initial segmentation step allowing for the isolation of the bones, and the outer linen layers is displayed in Figure 13. The three-point clouds generated from this first step were respectively constituted of 1,111,249; 1,451,332 and 7,567,975 points.

Immersive Volumetric Point Cloud Manipulation for Cultural Heritage

Rafik Drissi, DIGI ROOTS XR, Ronan Gaugne, Univ Rennes, Inria, CNRS, IRISA, France, Théophane Nicolas, Inrap, UMR 8215 Trajectoires, France, Valérie Gouranton, Univ Rennes, INSA Rennes, Inria, CNRS, IRISA, France

Focusing on the bones and making use of the scale features of the tool allows for the visualization of the bones at different scales as shown on Figure 14.
This allows for the user to experience "Walking" inside of a bone cavity and studying it at high detail.

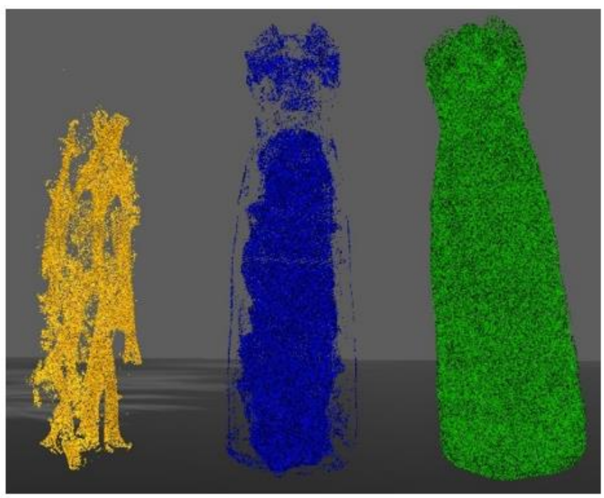


Figure 13: From left to right: Point clouds of the pre-segmented bones, dense wrappings and external wrappings

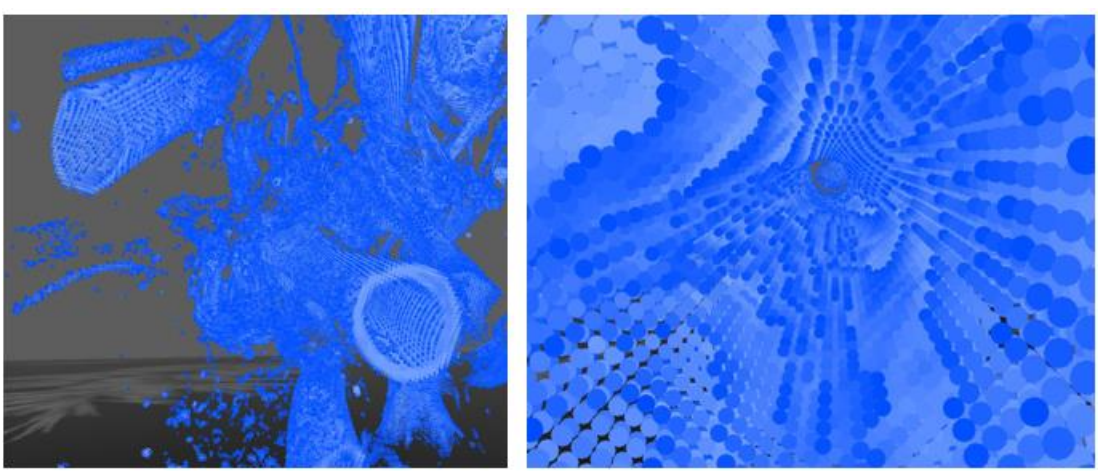


Figure 14: From left to right: Introspection of bones at initial size, and Introspection of oversized bone

Immersive Volumetric Point Cloud Manipulation for Cultural Heritage

Rafik Drissi, DIGI ROOTS XR, Ronan Gaugne, Univ Rennes, Inria, CNRS, IRISA, France, Théophane Nicolas, Inrap, UMR 8215 Trajectoires, France, Valérie Gouranton, Univ Rennes, INSA Rennes, Inria, CNRS, IRISA, France

Furthermore, the tool was used for real time extraction of specific bones for standalone analysis. The Figure 15 shows the extraction of the bone of a cat's paw from the cat's mummy, by moving the selection sphere (the gray ball)
Once segmented, the generated point cloud of the bone is easy to manipulate and visualize away from the neighboring elements and layers initially occluding it, allowing for the observation of the key elements for identification of the bones.

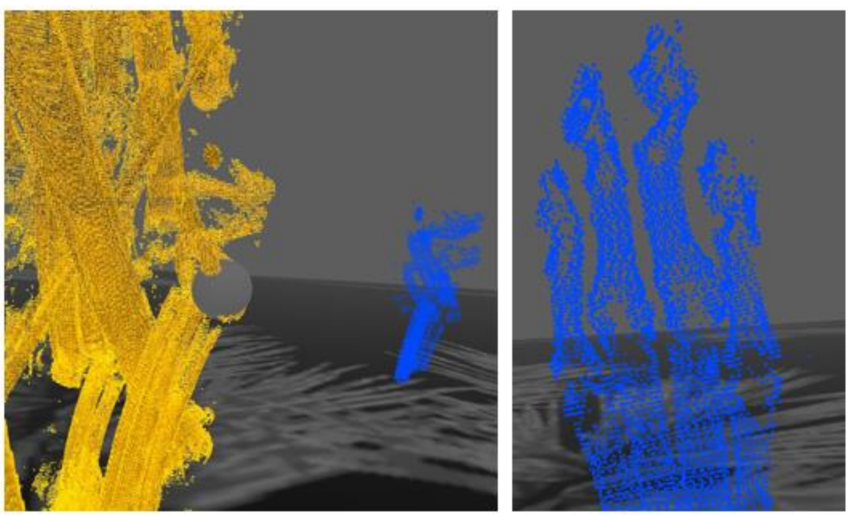


Figure 15: From left to right: Extraction of a cat's paw point cloud from the bone point cloud

C. Cagniard siren case :

The third archaeological material, the Cagnard Siren, was composed of several independently scanned objects, the socle, the mechanism and three whistles. All the parts were pre-segmented into two density thresholds:
The metal: density threshold from 2000 to 20 000
The wood: density threshold from -300 to 200
The conversion step resulted in the generation of 6 different point clouds whose sizes are presented in table 1.

Immersive Volumetric Point Cloud Manipulation for Cultural Heritage

Rafik Drissi, DIGI ROOTS XR, Ronan Gaugne, Univ Rennes, Inria, CNRS, IRISA, France, Théophane Nicolas, Inrap, UMR 8215 Trajectoires, France, Valérie Gouranton, Univ Rennes, INSA Rennes, Inria, CNRS, IRISA, France

The generated point clouds are laid out in a 3d scene to reproduce the real setting of the siren. This use case highlights the interest of the cross-section tool to visualize the assembly of the different elements before making selections and isolating the point clouds. It also illustrates the importance of contrasting point cloud colors in the visualization as shown in Figure 16.

	metal density	wood density
Socle	1 154 391	1 476 255
Mechanism	4 725 453	NP
Small whistle	NP	724 185
Medium whistle	5 422	1 967 199

Table 1. Number of points per generated point cloud from the Cagniard Siren

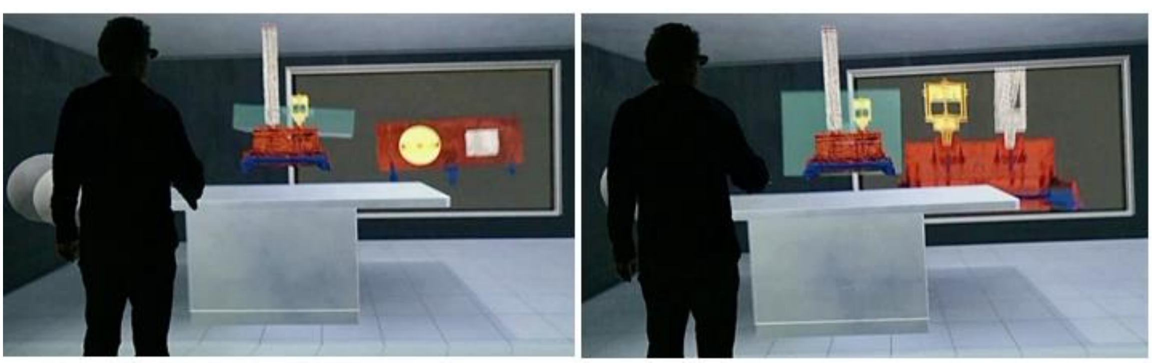


Figure 16: From Left to right: Horizontal cross section and vertical cross section of the siren point clouds

We used the framework's positional criteria at point cloud generation to generate a point cloud of the mechanism and a cross section point cloud revealing the internal gears as in Figure 17.

Immersive Volumetric Point Cloud Manipulation for Cultural Heritage

Rafik Drissi, DIGI ROOTS XR, Ronan Gaugne, Univ Rennes, Inria, CNRS, IRISA, France, Théophane Nicolas, Inrap, UMR 8215 Trajectoires, France, Valérie Gouranton, Univ Rennes, INSA Rennes, Inria, CNRS, IRISA, France

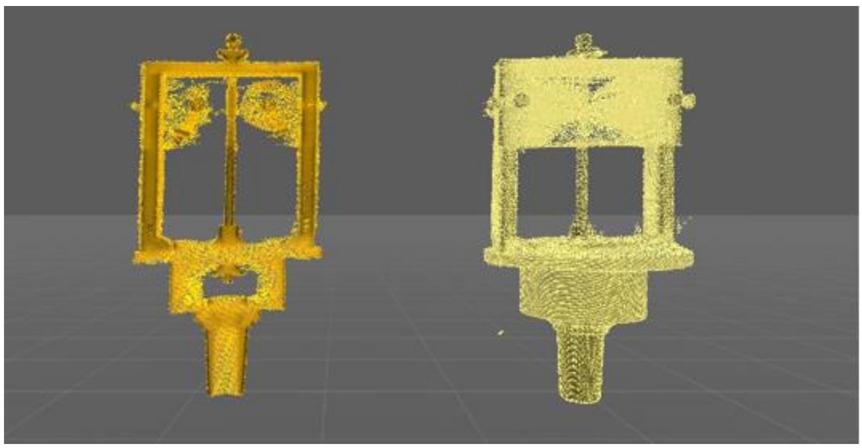


Figure 17: From left to right: point cloud of the cross section of the whistle, and point cloud of the entire whistle

The framework allows to visualize the composition of the studied object; to determine the nature of the components (metal, wood) and to identify the way these components were assembled during manufacture (internal screws and tubes). It is also possible to observe the usage of different types of wood and surface engravings as in the previous Figure 3.

VI. CONCLUSION

This paper presented a framework with a set of tools to manipulate volumetric point clouds in a highly immersive and interactive virtual reality environment, in a cultural heritage context. The tools were developed and used on different contexts and use cases. The framework offers an application for generation of the point clouds from the DICOM files, and an immersive application whose main functionalities are a segmentation tool that allows to extract any collection of points for independent manipulation, a scaling tool that allows for free roaming inside the point cloud and the ability to treat volume data from multiple sources in parallel for direct comparisons.

Immersive Volumetric Point Cloud Manipulation for Cultural Heritage

Rafik Drissi, DIGI ROOTS XR, Ronan Gaugne, Univ Rennes, Inria, CNRS, IRISA, France, Théophane Nicolas, Inrap, UMR 8215 Trajectoires, France, Valérie Gouranton, Univ Rennes, INSA Rennes, Inria, CNRS, IRISA, France

The increasing use of digitization in Cultural Heritage context with the production of more and more digital data justifies the development of adapted processing tools, especially based on point clouds which constitute a common volume data type in this field. Virtual reality, by immersing the user in these data is an interesting solution, especially by allowing to propose natural interactions to manipulate them. Further developments are underway to adapt the framework to a broader range of contexts. For example, standard tools such as measurement and annotations will be directly integrated as well as the support of simultaneous multiple users which will enable collaboration, as demonstrated in [5], as it allows archaeologists to share their respective expertise and to develop relevant hypotheses.

REFERENCES

[1] BACIM, F., NABIYOUNI, M., AND BOWMAN, D. A. Slice-n-swipe: A free-hand gesture user interface for 3d point cloud annotation. In 2014 IEEE Symposium on 3D User Interfaces (3DUI) (2014), pp. 185–186.

[2] CABRAL, M., MONTES, A., BELLOC, O., FERRAZ, R., TEUBL, F., DORETO, F., LOPES, R., AND ZUFFO, M. Bi-manual gesture interaction for 3d cloud point selection and annotation using cots. In 2014 IEEE Symposium on 3D User Interfaces (3DUI) (2014), pp. 187–188.

[3] CHOUGULE, V., MULAY, A., AND AHUJA, B. Three dimensional point cloud generations from ct scan images for bio-cad modeling.

[4] DROGEMULLER, A., CUNNINGHAM, A., WALSH, J., THOMAS, B. H., CORDEIL, M., AND ROSS, W. Examining virtual reality navigation techniques for 3d network visuali- sations. Journal of Computer Languages 56 (2020), 100937.

[5] FORTE, M., AND KURILLO, G. Cyberarchaeology: Exper- imenting with teleimmersive archaeology. In 2010 16th International Conference on Virtual Systems and Multimedia (2010), pp. 155–162.

[6] GAUGNE, R., PETIT, Q., BARREAU, J.-B., AND GOURAN-TON, V. Interactive and Immersive Tools for Point Clouds in Archaeology. In ICAT-EGVE 2019 - International Confer- ence on Artificial Reality and Telexistence and Eurographics Symposium on Virtual Environments (2019), Y. Kakehi and A. Hiyama, Eds., The Eurographics Association.

Immersive Volumetric Point Cloud Manipulation for Cultural Heritage

Rafik Drissi, DIGI ROOTS XR, Ronan Gaugne, Univ Rennes, Inria, CNRS, IRISA, France, Théophane Nicolas, Inrap, UMR 8215 Trajectoires, France, Valérie Gouranton, Univ Rennes, INSA Rennes, Inria, CNRS, IRISA, France

[7] HERMAN, G. T., AND LIU, H. K. Three-dimensional display of human organs from computed tomograms. Computer Graphics and Image Processing 9, 1 (1979), 1 – 21.

[8] HA¨ NEL, C., WEYERS, B., HENTSCHEL, B., AND KUHLEN, T. W. Interactive volume rendering for immersive virtual environments. In 2014 IEEE VIS International Workshop on 3DVis (3DVis) (2014), pp. 73–74.

[9] JANSEN, R. J., POULUS, M., KOTTMAN, J., DE GROOT, T., HUISMAN, D. J., AND STOKER, J. Ct: a new nondestructive method for visualizing and characterizing ancient roman glass fragments in situ in blocks of soil. Radiographics 26, 6 (2006), 1837–1844.

[10] KRAMMES, H., SILVA, M. M., MOTA, T., TURA, M. T., MACIEL, A., AND NEDEL, L. P. The point walker multi- label approach. In IEEE Symposium on 3D User Interfaces, 3DUI 2014, Minneapolis, MN, USA, March 29-30, 2014 (2014), A. Le´cuyer, R. Lindeman, and F. Steinicke, Eds., IEEE Computer Society, pp. 189–190.

[11] KREYLOS, O., BAWDEN, G. W., AND KELLOGG, L. H. Immersive visualization and analysis of lidar data. In Advances in Visual Computing (2008), G. Bebis, R. Boyle, B. Parvin, and al., Eds., Springer Berlin Heidelberg, pp. 846–855.

[12] KULIK, A., KUNERT, A., BECK, S., MATTHES, C.-F., SCHOLLMEYER, A., KRESKOWSKI, A., FRO¨ HLICH, B., COBB, S., AND D'CRUZ, M. Virtual valcamonica: Collaborative exploration of prehistoric petroglyphs and their surrounding environment in multi-user virtual reality. Presence: Teleoperators and Virtual Environments 26 (2018).

[13] MALOCA, P., DE CARVALHO, E., HEEREN, T., HASLER, P., MUSHTAQ, F., MON-WILLIAMS, M., SCHOLL, H., BAL- ASKAS, K., EGAN, C., TUFAIL, A., WITTHAUER, L., AND CATTIN, P. High-performance virtual reality volume rendering of original optical coherence tomography point-cloud data enhanced with real-time ray casting. Translational Vision Science & Technology 7 (07 2018).

[14] MONTANO-MURILLO, R. A., NGUYEN, C., KAZI, R. H., SUBRAMANIAN, S., DIVERDI, S., AND MARTINEZ-PLASENCIA, D. Slicing-volume: Hybrid 3d/2d multi-target selection technique for dense virtual environments. In 2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR) (2020), pp. 53–62. NICOLAS, T., GAUGNE, R., TAVERNIER, C., MILLET, E., BERNADET, R., AND GOURANTON, V. Lift the veil of the block samples from the Warcq chariot burial with 3D digital technologies. In Digital Heritage 2018 - 3rd International Congress & Expo, IEEE (San Francisco, United States, 2018), pp. 1–8.

[15] PALHA, A., MURTIYOSO, A., MICHELIN, J.-C., ALBY, E., AND GRUSSENMEYER, P. Open source first person view 3d point cloud visualizer for large data sets. In The Rise of Big Spatial Data (Cham, 2017), I. Ivan, A. Singleton, J. Hora´k, and T. Inspektor, Eds., Springer International Publishing, pp. 27–39.

Immersive Volumetric Point Cloud Manipulation for Cultural Heritage

Rafik Drissi, DIGI ROOTS XR, Ronan Gaugne, Univ Rennes, Inria, CNRS, IRISA, France, Théophane Nicolas, Inrap, UMR 8215 Trajectoires, France, Valérie Gouranton, Univ Rennes, INSA Rennes, Inria, CNRS, IRISA, France

[16] RE, A., CORSI, J., DEMMELBAUER, M., MARTINI, M., MILA, G., AND RICCI, C. X-ray tomography of a soil block: A useful tool for the restoration of archaeological finds. Heritage Science 3, 4 (01 2015).

[17] STELZNER, J., EBINGER-RIST, N., PEEK, C., AND SCHILLINGER, B. The application of 3d computed tomography with x-rays and neutrons to visualize archaeological objects in blocks of soil. Studies in Conservation 55 (06 2010), 95–106.

[18] VEIT, M., AND CAPOBIANCO, A. Go'then'tag: A 3-d point cloud annotation technique. In 2014 IEEE Symposium on 3D User Interfaces (3DUI) (2014), pp. 193–194.

Visual Computing Magazine

A Systematic Procedure for Comparing Template-based Gesture Recognizers

Mehdi Ousmer, LRIM, University Catholic of Louvain La Neuve, Belgium

Abstract
This paper presents a conceptual framework and systematic procedure for conducting comparative testing of template-based gesture recognizers under controlled conditions. The proposed systematic procedure investigates the influence of various factors on recognition accuracy and responsiveness, including the number of templates, sampling points, fingers, and their configuration with other hand parameters, such as the palm. The experimental validation applies 6 selected gesture recognition algorithms tested on 2 established benchmark datasets from the literature and gesture recognition competitions.

1 Introduction

A diverse array of input devices, including game controllers, smartphones, and optical sensors, can be incorporated into three-dimensional gesture-based user interfaces (3D GBUI). These devices typically incorporate multiple motion sensors capable of measuring various physical quantities, thereby facilitating the broad application of gesture recognition algorithms in multiple domains.

This article focuses on designing and applying a systematic procedure for comparative testing. We rely on an adapted version of the AMMPT framework throughout this work for consistency in terminology [2, 17]. A systematic procedure refers to a structured sequence of steps followed in a defined order, while comparative testing evaluates software performance against competitors to improve quality and functionality [25]. While this method offers benefits for software improvement and market positioning, it can face challenges in implementing changes and managing customer perceptions [1].

Despite extensive comparative analyses in gesture recognition research [10,11], a comprehensive evaluation of hand gesture recognition using the Leap motion controller remains unexplored.

Extensive research exists on gesture recognition using advanced Machine Learning (ML) techniques for LMC [8]. These models excel at recognizing complex gestures and handling noisy data [9, 21]. Deep learning-based recognizers have become particularly prevalent [4, 20]. However, these techniques require large gesture datasets for training and validation to achieve accurate results with complex gestures.

Visual Computing Magazine

A Systematic Procedure for Comparing Template-based Gesture Recognizers

Mehdi Ousmer, LRIM, University Catholic of Louvain La Neuve, Belgium

This paper focuses on template matching recognizers. This technique provides a fast, simple, and accurate classification method. A template matching recognizer identifies candidate gestures by comparing them with predefined gesture classes, which are represented by a limited set of training templates. The selection of this technique is justified by several criteria:

1. Real-time interaction requirements for interactive applications demand optimal system performance, characterized by minimal response time (e.g., 0.1 seconds, according to Nielsen [15]) and high accuracy (e.g., a recognition rate of >90% [14, 26]).
2. A gesture dataset that's easy to design, modify, and expand, particularly for supporting user-defined gestures [27].
3. They require minimal AI/ML expertise, as they are easy to train and modify, understand and compute, and interpret, and offer straightforward integration into the development cycle.

2 Methodology

2.1 Gesture Recognition Testing

The gesture recognition testing procedure includes four distinct stages: elicitation, data acquisition, training, and testing. Although these stages constitute standard practice in the literature on gesture recognition evaluation, their implementation varies between works [13, 12, 28].

The initial phase, the Gesture Elicitation Study, enables collaborative gesture creation wherein participants design task-specific gestures reflecting end user preferences. In the subsequent data acquisition stage, multiple devices are used to capture gesture data. The training stage focuses on the development of the recognizer model based on the selected classification technique.

The final testing stage evaluates the trained model's recognition capabilities through performance metrics. To formally represent the gesture recognition testing procedure, the Software & Systems Process Engineering Meta-Model (SPEM) notation was employed [16, 17]) (Figure 1).

A Systematic Procedure for Comparing Template-based Gesture Recognizers

Mehdi Ousmer, LRIM, University Catholic of Louvain La Neuve, Belgium

2.2 Comparative Testing

Gesture recognition comparative testing evaluates template-based recognizers' performance in multiple contexts of use. The testing usually involves comparing the accuracy or recognition rate of the recognizers and considering factors such as computational complexity, real-time capability, and feasibility on affordable hardware. Comparative testing involves evaluating recognizers against predefined test cases to identify the most accurate one. This process relies on a structured framework for comparing different template-based gesture recognizers. The literature identifies several key components essential for gesture recognition testing.

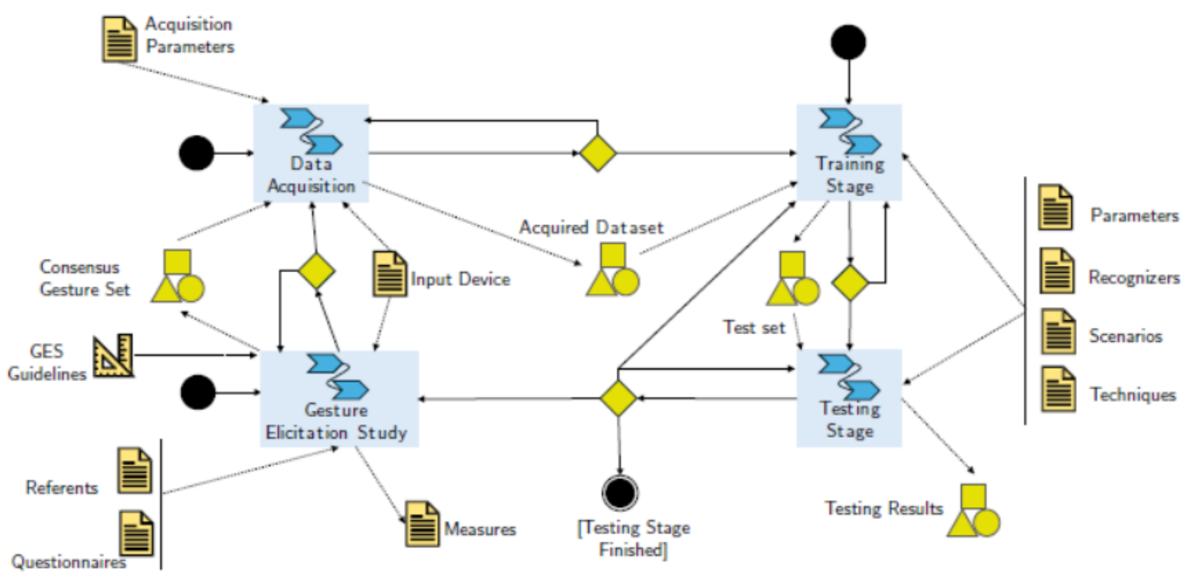


Figure 1: Gesture recognition testing procedure workflow. [17]

A Systematic Procedure for Comparing Template-based Gesture Recognizers

Mehdi Ousmer, LRIM, University Catholic of Louvain La Neuve, Belgium

2.2.1 Recognizers
Template-based recognizers form the core testing component. They identify gestures by matching them against training templates. These recognizers must use the same programming language and be compatible with 3D hand gesture datasets.

2.2.2 Techniques
Many key evaluation techniques exist. These include k-fold cross-validation (dividing data into k test groups), 50-50 splits, leave-one-out cross-subject evaluation, and leave-one-out cross-session evaluation[5].

2.2.3 Scenarios
Testing covers single and multi-device configurations. Two main approaches are used: user-dependent (same user for training and testing) and user-independent (separate users for training and testing).

2.2.4 Datasets
Testing relies on gesture datasets collected from various users and devices. Public datasets are preferred for reproducibility. Dataset selection must consider the context of use, gesture diversity and variability, and data availability.

2.2.5 Parameters
The main parameters include the number of templates (T) and repetitions (number of tests) (R) for the user-dependent scenario, plus the number of participants (P) for the user-independent scenario, with additional parameters for multi-device scenarios. Parameter values are based on experience, prior research, and dataset constraints.

2.3 Leap Motion Comparative Testing

The gesture recognition testing procedure shares elements with comparative testing. A unified testing process with repeatable stages that provides a systematic approach to this procedure through clearly defined activities. The use of consistent activities ensures completeness and systematization. This standardization enables valid comparisons between recognition algorithms, while using different procedures would lead to incomparable results.

Visual Computing Magazine

A Systematic Procedure for Comparing Template-based Gesture Recognizers

Mehdi Ousmer, LRIM, University Catholic of Louvain La Neuve, Belgium

The comparative testing presented in the following section aims to fill the gap in the literature [19]. We selected many datasets that were available for various reasons, including reproducibility. We evaluate the recognizers ($P3+ [18],$F [18], $P3+X: a variant of $P3+, Jackknife [23], PennyPincher3D: a 3D adaptation of [22], 3¢ [7]) in a user-independent scenario. We test them on the full hand gestures provided by the LMC (Leap Motion Controller) skeletal hand model available in the SHREC2019 [6] (Figure 2) and Jackknife-LM [23] datasets (Figure 3).

We measure both recognition rate and execution time for the 12 basic configurations (6 RECOGNIZER × 2 DATASET), following established evaluation methods used in the literature to evaluate gesture recognizers [3, 27, 24]: the user-independent scenario evaluates the recognition on gestures produced by users who are different from those used for training the recognizer. In this scenario, the basic configurations are refined depending on A (Figure 4), the number of joints, on T, the number of templates, and depending on N, the number of resampling points to train the recognizer.

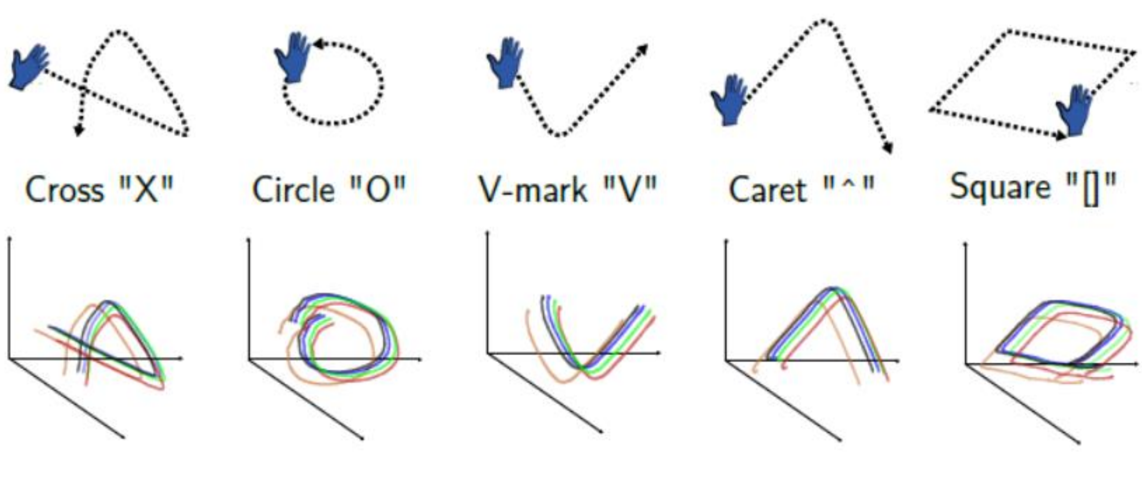


Figure 2: The SHREC2019 gesture classes and samples. [6]

Visual Computing Magazine

A Systematic Procedure for Comparing Template-based Gesture Recognizers

Mehdi Ousmer, LRIM, University Catholic of Louvain La Neuve, Belgium

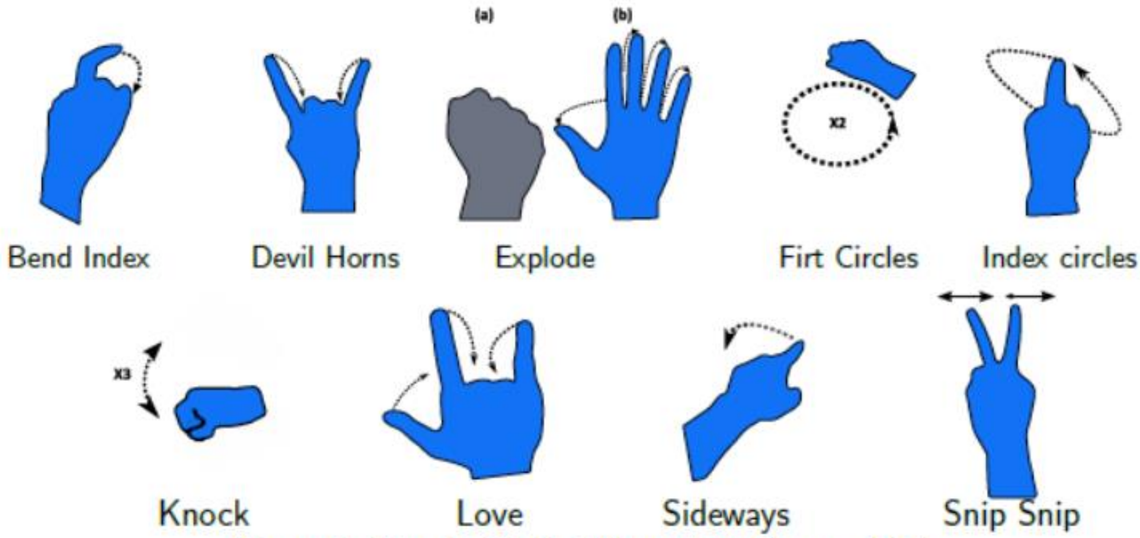


Figure 3: The Jackknife-LM gesture classes. [23]

3 Results & discussion

3.1 SHREC2019 Overall Recognition Rate

Figure 5 shows the average recognition rates on all tests. $P3+ leads with (M=85.90%, SD=15.44%), followed by $P3+X (M=84.90%, SD=16.45%) and Jackknife (M=82.11%, SD=14.06%). PennyPincher3D achieves (M=80.02%, SD=12.38%), while 3 Cent (M=78.09%, SD=18.12%) and $F (M=77.36%, SD=19.27%) fall below 80%.

Statistical analysis revealed that none of the recognition rates followed a normal distribution on the RECOGNIZER variable. Kruskal-Wallis testing showed highly significant differences between recognizers (p<.001***). $P3+ significantly outperformed all others, including $F (Z=48.84, p<.001***), Jackknife (Z=27.86, p<.001***), and $P3+X (Z=5.247, p<.001***). While $F and 3 Cent showed no significant difference (Z=1.928, n.s.), PennnyPincher3D outperformed 3 Cent (Z=3.055, p<.05*). For specific gesture classes:

- “Caret” and “V-mark” have average recognition rates greater than or equal to 90%.
- “Cross” outperformed “Square” and “Circle” in $F, $P3+ and $P3 + X.

A Systematic Procedure for Comparing Template-based Gesture Recognizers

Mehdi Ousmer, LRIM, University Catholic of Louvain La Neuve, Belgium

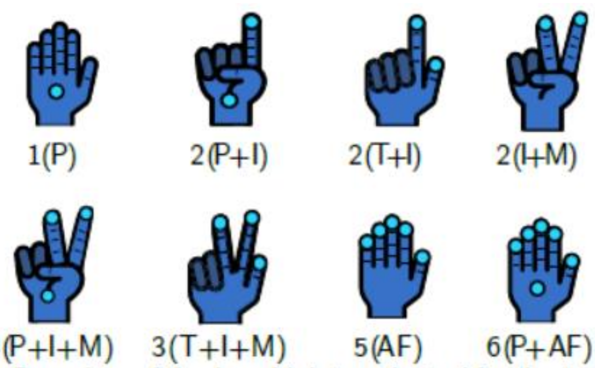


Figure 4: Overview of the hand joints selected for the testing [*A*].

• Jackknife best recognized “Cross” (M=93.98%), compared to $P3+ (M=83.35%) and $F (M=67.75%).
• “Square” and “Cross” showed similar recognition rates for Jackknife and PennyPincher3D.
• “Circle” performed poorly with Jackknife (M=34.43%), PennyPincher3D (M=21.91%), and 3 Cent (M=46.50%).

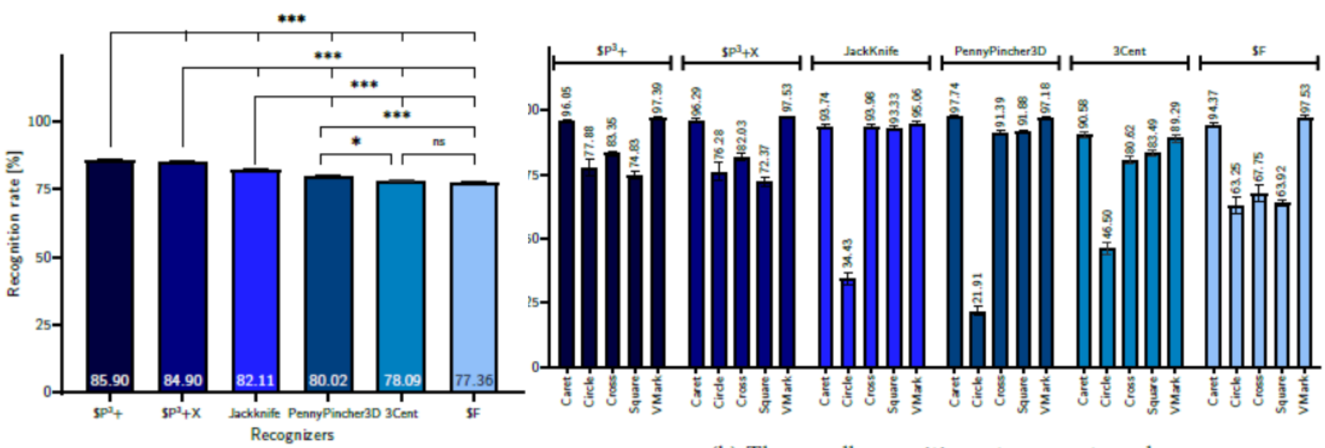
While PennyPincher3D, 3 Cent, and $F rarely met the 90% user expectation [14, 26], $P3+ and $P3 +X consistently did. For SHREC2019 testing, we relaxed this to τ≥80%.

3.2 SHREC2019 Overall Execution Time

In Figure 6, recognizer execution times show that, the PennyPincher3D is the fastest (M=0.047 ms, SD=0.060 ms), while the $P3+ (M=1.611 ms, SD=3.328ms) is the slowest recognizer. Kruskal-Wallis and Dunn’s tests revealed significant differences between all recognizer pairs except 3 Cent and $P3+X. PennyPincher3D was significantly faster than Jackknife (M=0.195 ms, difference: 148μs, Z=73.62, p<.001) and $P3+ (Z=147.7, p<.001).

A Systematic Procedure for Comparing Template-based Gesture Recognizers

Mehdi Ousmer, LRIM, University Catholic of Louvain La Neuve, Belgium



(a) Average recognition Rate

(b) The overall recognition rate per gesture class

Figure 5: Recognition rates of all recognizers for the SHREC2019 for all conditions. Error bars show a confidence interval of 95%.

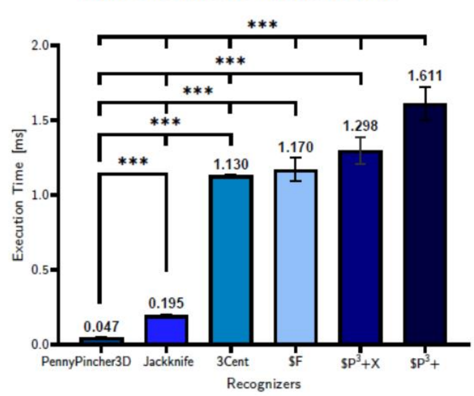


Figure 6: Average execution times by recognizer. Error bars show a confidence interval of 95%.

3.3 Jackknife-LM Overall Recognition Rate

The Jackknife-LM dataset contains more complex gesture classes than SHREC2019, where different joints can move independently of the hand movement. Testing results in Figure 7 showed no recognizer exceeded 80% average accuracy for all tests under all conditions. Jackknife performed best (M=73.60%, SD=19.75%), followed by $P3+ (M=68.75%, SD=19.73%) and $P3+X (M=68.15%, SD=20.51%).

A Systematic Procedure for Comparing Template-based Gesture Recognizers

Mehdi Ousmer, LRIM, University Catholic of Louvain La Neuve, Belgium

Lower performance was seen from 3 Cent (M=63.66%, SD=18.29%), $F (M=62.96%, SD=20.12%), and PennyPincher3D (M=56.87%, SD=19.99%). Most recognizers achieved high accuracy (τ≥80%) for “FistCircles”, “Knock” and “Sideways” gestures, except PennyPincher3D which scored lower (M=62.77%) for “Knock”. “Love”, “BendIndex”, “DevilHorns”, and “SnipSnip” showed poor recognition (τ≤60%) in most recognizers, with Jackknife performing better (> 70%) on “SnipSnip” and “BendIndex”. Results show better recognition for static finger poses compared to gestures involving finger movements.

3.4 Jackknife-LM Overall Execution Time

The average execution time results in Figure 8 showed significant variations between recognizers. PennyPincher3D was fastest (M=0.079 ms, SD=0.108 ms), while 3 Cent was slowest (M=10.585 ms, SD=5.911 ms) due to its Cubic Spline interpolation. $P3+, $F, and $P3 +X had similar performance (around 2 ms), while Jackknife combined high accuracy with fast execution (M=0.351ms, SD=0.508 ms).

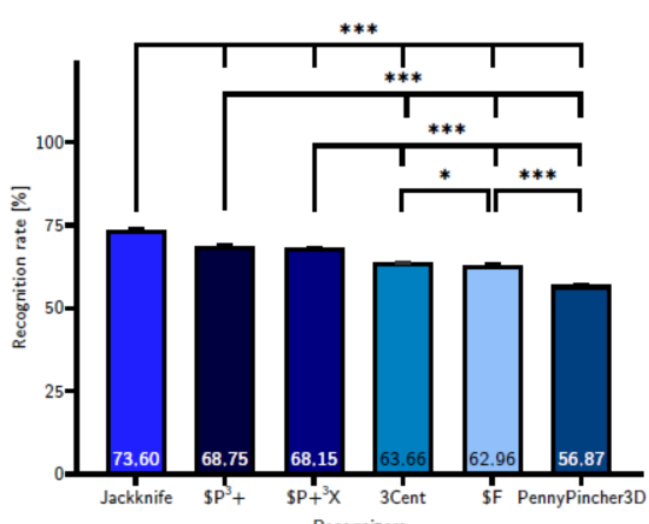


Figure 7: Recognition rates of all recognizers for the Jackknife-LM for all conditions. Error bars show a confidence interval of 95%.

A Systematic Procedure for Comparing Template-based Gesture Recognizers

Mehdi Ousmer, LRIM, University Catholic of Louvain La Neuve, Belgium

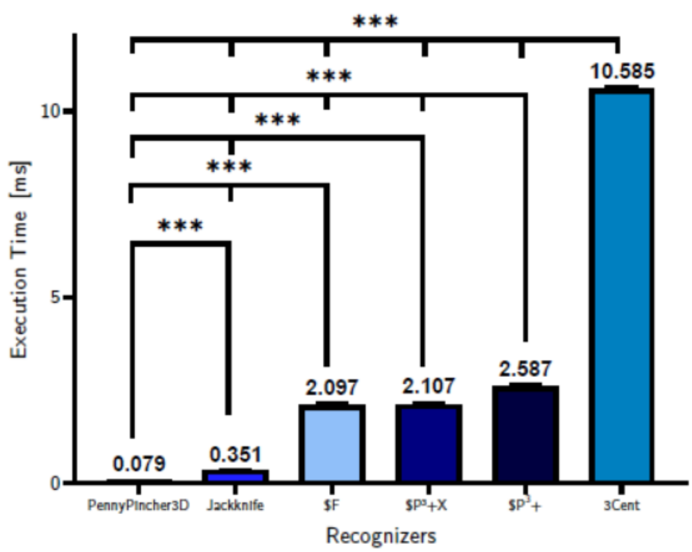


Figure 8: Average execution times by recognizer. Error bars show a confidence interval of 95%.

4 Conclusion

The comparative testing of the six selected gesture recognizers demonstrates the potential benefits of the systematic procedure for the comparative testing method. It provides many significant results depending on the dataset: the SHREC2019 dataset (an LMC-Based dataset with simple gestures) and the Jackknife-LM (which contains more complex gestures). Overall, the evaluation results showed that the recognizers perform differently for each dataset.

We analyzed the recognizers under many conditions and compared them under particular optimal conditions [19, 17]. These tests confirmed some of our previous results, showing that certain recognizers perform well for several configurations. Furthermore, we found that some recognizers are better suited to recognize specific gestures.

A Systematic Procedure for Comparing Template-based Gesture Recognizers

Mehdi Ousmer, LRIM, University Catholic of Louvain La Neuve, Belgium

References

[1] A. Ahmad. What is comparison testing, 2022.
[2] V. Andiappan and Y.Wan. Distinguishing approach, methodology, method, procedure and technique in process systems engineering. Clean Technologies and Environmental Policy, 22(3):547–555, apr 2020.
[3] L. Anthony and J. O. Wobbrock. A lightweight multistroke recognizer for user interface prototypes. In Proceedings of Graphics Interface 2010, GI '10, pages 245–252, Toronto, Ont., Canada, Canada, 2010. Canadian Information Processing Society.
[4] M. Asadi-Aghbolaghi, A. Clap´es, M. Bellantonio, H. J. Escalante, V. Ponce-L´opez, X. Bar´ o, I. Guyon, S. Kasaei, and S. Escalera. Deep Learning for Action and Gesture Recognition in Image Sequences: A Survey, pages 539–578. The Springer Series on Challenges in Machine Learning. Springer International Publishing, Cham, 2017.
[5] A. D. Berenguer, M. C. Oveneke, H.-U.-R. Khalid, M. Alioscha-Perez, A. Bourdoux, and H. Sahli. Gesturevlad: Combining unsupervised features representation and spatio-temporal aggregation for doppler-radar gesture recognition. IEEE Access, 7:137122–137135, 2019.
[6] F. M. Caputo, S. Burato, G. Pavan, T. Voillemin, H. Wannous, J. P. Vandeborre, M. Maghoumi, E. M. Taranta II, A. Razmjoo, J. J. LaViola Jr., F. Manganaro, S. Pini, G. Borghi, R. Vezzani, R. Cucchiara, H. Nguyen, M. T. Tran, and A. Giachetti. Online Gesture Recognition. In S. Biasotti, G. Lavou´ e, and R. Veltkamp, editors, Eurographics Workshop on 3D Object Retrieval, pages 93–102. The Eurographics Association, 2019.
[7] F. M. Caputo, P. Prebianca, A. Carcangiu, L. D. Spano, and A. Giachetti. A 3 cent recognizer: Simple and effective retrieval and classification of mid-air gestures from single 3d traces. In Proceedings of the Conference on Smart Tools and Applications in Computer Graphics, STAG '17, page 9–15, Goslar, DEU, 2017. Eurographics Association.
[8] I. A. S. Filho, E. N. Chen, J. M. da Silva Junior, and R. da Silva Barboza. Gesture recognition using leap motion: A comparison between machine learning algorithms. In ACM SIGGRAPH 2018 Posters, SIGGRAPH '18, New York, NY, USA, 2018. Association for Computing Machinery.
[9] H. M. Jais, Z. R. Mahayuddin, and H. Arshad. A review on gesture recognition using kinect. In 2015 International Conference on Electrical Engineering and Informatics (ICEEI), page 594–599, Denpasar, Bali, Indonesia, aug 2015. IEEE.
[10] A. S. Khalaf, S. A. Alharthi, I. Dolgov, and Z. O. Toups. A comparative study of hand gesture recognition devices in the context of game design. In Proceedings of the 2019 ACM International Conference on Interactive Surfaces and Spaces, ISS '19, page 397–402, New York, NY, USA, 2019. Association for Computing Machinery.
[11] R. Z. Khan. Comparative study of hand gesture recognition system. Volume 2, pages 203–213, 07 2012.
[12] S. Kotsiantis, I. Zaharakis, and P. Pintelas. Machine learning: A review of classification and combining techniques. Artificial Intelligence Review, 26:159–190, nov 2006.
[13] J. J. LaViola. 3d gestural interaction: The state of the field. International Scholarly Research Notices, 2013, 2013.

A Systematic Procedure for Comparing Template-based Gesture Recognizers

Mehdi Ousmer, LRIM, University Catholic of Louvain La Neuve, Belgium

[14] G. Marin, F. Dominio, and P. Zanuttigh. Hand gesture recognition with jointly calibrated leap motion and depth sensor. Multimedia Tools Appl., 75(22):14991–15015, November 2016.
[15] J. Nielsen. Usability Engineering. Interactive Technologies. Elsevier Science, 1994.
[16] O. M. G. (OMG). Software & systems process engineering metamodel specification, version 2.0. http://www.omg.org/spec/SPEM/2.0/PDF/, Apr 2008.
[17] M. Ousmer. A Systematic Procedure for Comparative Testing of 3D Hand Gesture Template-Based Recognizers in Multiple Contexts of Use. PhD thesis, UCL UCL UCL - SSH/LouRIM - Louvain Research Institute in Management and Organizations SST/ICTM - Institute of Information and Communication Technologies, Electronics and Applied Mathematics Ecole Polytechnique de Louvain, 2024.
[18] M. Ousmer, A. Slu¨yters, N. Magrofuoco, P. Roselli, and J. Vanderdonckt. Recognizing 3d trajectories as 2d multi-stroke gestures. Proceedings of the ACM on Human-Computer Interaction, 4(ISS):1–21, nov 2020.
[19] M. Ousmer, A. Slu¨yters, N. Magrofuoco, P. Roselli, and J. Vanderdonckt. A systematic procedure for comparing template-based gesture recognizers. In M. Kurosu, S. Yamamoto, H. Mori, D. D. Schmorrow, C. M. Fidopiastis, N. A. Streitz, and S. Konomi, editors, HCI International 2022 - Late Breaking Papers. Multimodality in Advanced Interaction Environments, pages 160–179, Cham, 2022. Springer Nature Switzerland.
[20] S. Sharma and S. Singh. Vision-based hand gesture recognition using deep learning for the interpretation of sign language. Expert Systems with Applications, 182:115657, 2021.
[21] J. Suarez and R. R. Murphy. Hand gesture recognition with depth images: A review. In 2012 IEEE RO-MAN: The 21st IEEE International Symposium on Robot and Human Interactive Communication, page 411–417. IEEE,sep 2012.
[22] E. M. Taranta, II and J. J. LaViola, Jr. Penny pincher: A blazing fast, highly accurate $-family recognizer. In Proceedings of the 41st Graphics Interface Conference, GI '15, pages 195–202, Toronto, Ont., Canada, Canada, 2015. Canadian Information Processing Society.
[23] E. M. Taranta II, A. Samiei, M. Maghoumi, P. Khaloo, C. R. Pittman, and J. J. LaViola Jr. Jackknife: A reliable recognizer with few samples and many modalities. In Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, CHI '17, pages 5850–5861, New York, NY, USA, 2017. ACM.
[24] R.-D. Vatavu, L. Anthony, and J. O. Wobbrock. Gestures as point clouds: A $p recognizer for user interface prototypes. In Proceedings of the 14th ACM International Conference on Multimodal Interaction, ICMI '12, pages 273–280, New York, NY, USA, 2012. ACM.
[25] T. Vineet Nanda. Comparison testing in software engineering, 2021.
[26] C. Wang, Z. Liu, and S.-C. Chan. Superpixel-based hand gesture recognition with kinect depth camera. IEEE Transactions on Multimedia, 17(1):29–39, 2015.
[27] J. O. Wobbrock, A. D. Wilson, and Y. Li. Gestures without libraries, toolkits or training: A $1 recognizer for user interface prototypes. In Proceedings of the 20th Annual ACM Symposium on User Interface Software and Technology, UIST '07, pages 159–168, New York, NY, USA, 2007. ACM.
[28] M. Yasen and S. Jusoh. A systematic review on hand gesture recognition techniques, challenges and applications. PeerJ Computer Science, 5:e218, September 2019.

Visual Computing Magazine

Deep Learning-Based Hand Pose Estimation: A Novel ResUnet Framework for Occlusion Handling in AR/VR Applications

Roumaissa BEKIRI, Mohamed Chaouki BABAHENINI
LESIA Laboratory, Mohamed Khider University, Biskra, Algeria.

Abstract

Hand pose estimation is crucial for immersive human-computer interaction in augmented and virtual reality (AR/VR) applications, but occlusion remains a significant challenge. This paper presents ResUnet, a new deep learning framework that combines ResNet-34 with the U-Net architecture to estimate 2D and 3D hand poses from single RGB images while considering occlusion. Extensive experiments on three reference datasets (GANerated, SynthHands and STB) demonstrate performance that improves on existing techniques by 8-20%. Qualitative results validate the model's ability to reconstruct accurate 3D self-occlusion poses.

Keywords: Hand pose estimation, Deep learning, Occlusion handling, ResUnet, 3D pose regression.

1 Introduction

Estimating hands has become a long-standing research area in computer vision with several applications, including creating virtual and augmented reality systems and Mixed Reality(MR). Many powerful devices were invented in AR/VR/MR technology, such as Microsoft Hololens, HTC VIVE, or PlayStation VR, thus enabling a plausible interaction between human hands and virtual objects.
Estimating the hand is an essential part of many occasions that has garnered a great deal of popularity due to its effectiveness and versatility, which can be employed for gesture recognition in a variety of situations, as indicated by Piumsomboon et al.[1] the results of a guessability on hand gesture in AR environments in 40 selected tasks with 800 gestures. Yin et al.[2] created a system based on hand pose estimation to understand sign languages. As Markusseen et al.[3] defined a mid-air interaction using a keyboard to type in the air. Jang et al.[4] developed an AR / VR application to estimate the location of the hand in an egocentric perspective using the user's hands to control objects, as shown in Figure 1.
During 2015–2019, there was a significant increase in interest in estimating hand poses, focusing on deep learning methodologies, prominently featuring CNN.

Visual Computing Magazine

Deep Learning-Based Hand Pose Estimation: A Novel ResUnet Framework for Occlusion Handling in AR/VR Applications

Roumaissa BEKIRI, Mohamed Chaouki BABAHENINI
LESIA Laboratory, Mohamed Khider University, Biskra, Algeria.

This trend is supported by the findings of approximately 60 surveys, as highlighted in the work by Le et al. in 2020[5]. So, we conducted a new taxonomy based on deep learning in the hand pose estimation field, which is grouped into three categories depending on the input data. We identified three categories in our categorization: depth-based, image-based, and RGBD-based.

2 Methodology

Our main objective is to propose a novel deep learning architecture to estimate 2D and 3D poses from a single RGB image, destined to solve the occlusion problem.

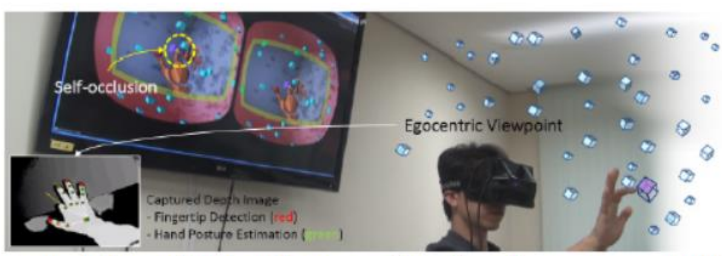


Fig. 1: A framework for using hand pose estimation from an egocentric perspective. Initially applied in[4]

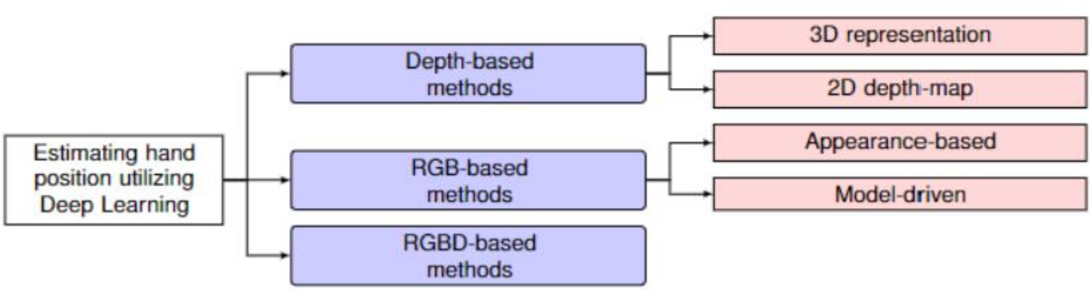


Fig. 2: A proposed taxonomy for hand pose estimation using deep learning

Deep Learning-Based Hand Pose Estimation: A Novel ResUnet Framework for Occlusion Handling in AR/VR Applications

Roumaissa BEKIRI, Mohamed Chaouki BABAHENINI
LESIA Laboratory, Mohamed Khider University, Biskra, Algeria.

The 3D hand pose is represented by a sequence of 3D joint coordinates, $\Phi^{3D} = \{\phi\}_{k=1}^{K} \in T_{3D}$ where T3D is the location of the 3D hand joint, with K = 21. The 2D hand pose estimation is represented by a coordinated two-dimensional array joint, where is the s$\Phi^{2D} = \{\phi\}_{k=1}^{K} \in S_{2D}$hand joint with K = 21.

The proposed ”ResUnet” framework combines ResNet-34 layers with Unet as its foundational network. This combination is particularly effective for tasks where input and output have similar dimensions. The Unet network consists of two primary pathways. The first path utilizes a pre-trained ResNet-34, which is a 34-layer ResNet network, to extract key features from an RGB cropped hand image denoted as
$I \in +R^{128\times128\times3}$, The second path of Unet, known as the Expansive path, includes four consecutive multi-feature blocks combined with upsample Blocks referred to as Unet-Blocks, as illustrated in Figure 3. The initial Unet-Block takes two features, denoted as Fg = F4, F3, fuses them as input, and produces grouped features denoted as Fout. We employ bilinear upsampling to enhance the quality of the input images and obtain multi-scale features. Subsequently, these features are used to estimate 2D heat maps before passing them on to the next Unet-Block.
The following Unet-Blocks follow a similar structure but use distinct input features. The upsampling layer is concatenated for each block with the corresponding feature vector, denoted as F_{skip}. These concatenated features are later fed into the convolutional layers for further processing.

3 Experimental Evaluation and Results

We have conducted quantitative and qualitative evaluations of our technique to address the challenge of learning 2D and 3D hand pose regression, explicitly targeting the issue of occlusion.

Deep Learning-Based Hand Pose Estimation: A Novel ResUnet Framework for Occlusion Handling in AR/VR Applications

Roumaissa BEKIRI, Mohamed Chaouki BABAHENINI
LESIA Laboratory, Mohamed Khider University, Biskra, Algeria.

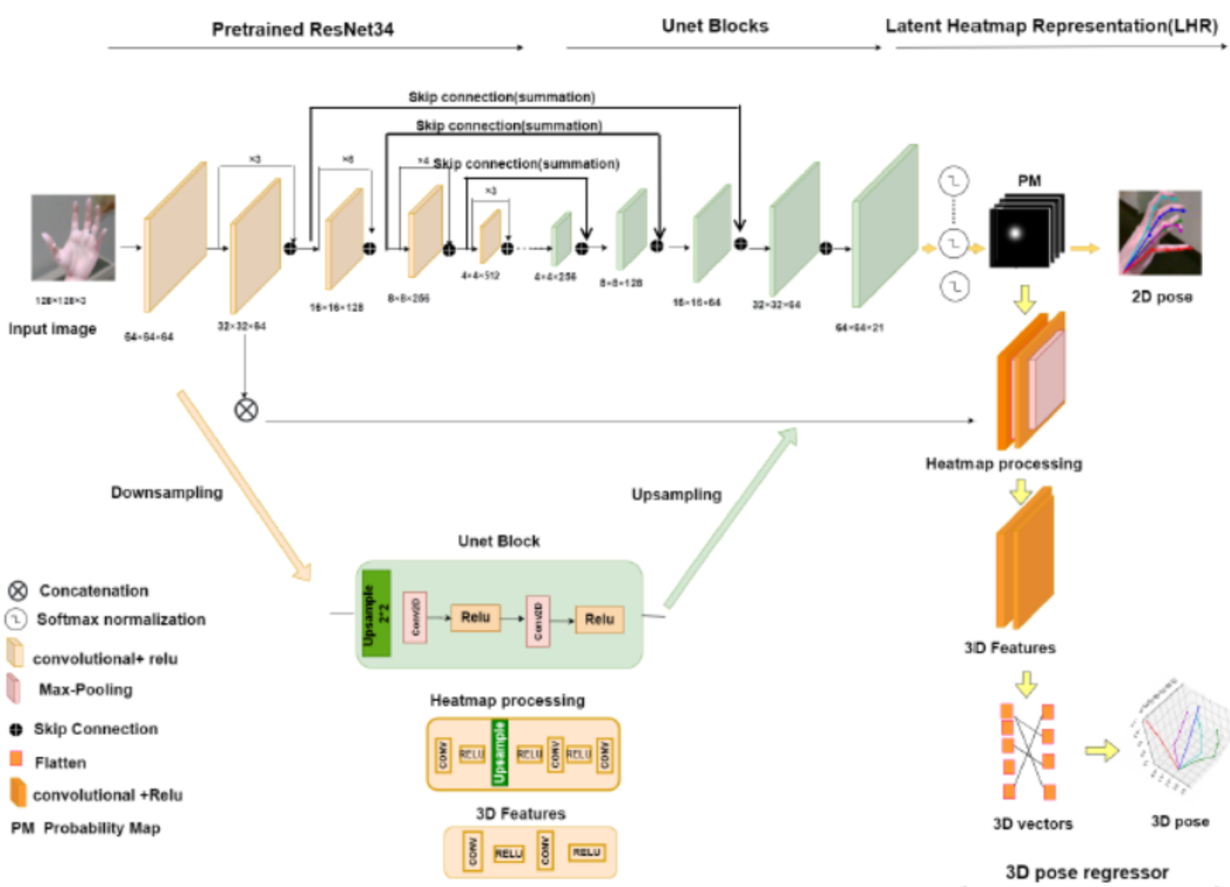


Fig. 3: Overall pipeline of our architecture for estimating 2D and 3D pose regression.

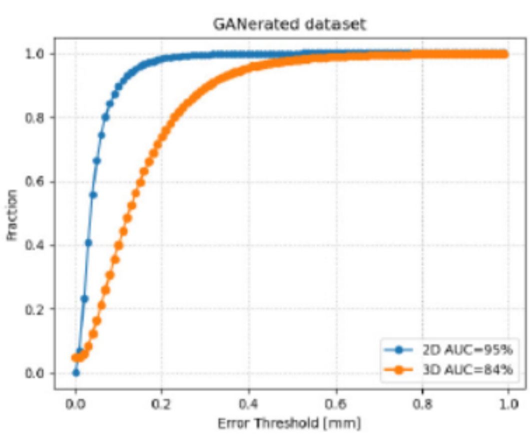
3.1 Quantitative evaluation

To quantitatively validate our results and enhance the robustness of our strategy, we have employed various metrics for quality assessment. We evaluated the robustness of our architecture across three datasets: the GANerated dataset, the Stereo dataset, and the SynthHands dataset.

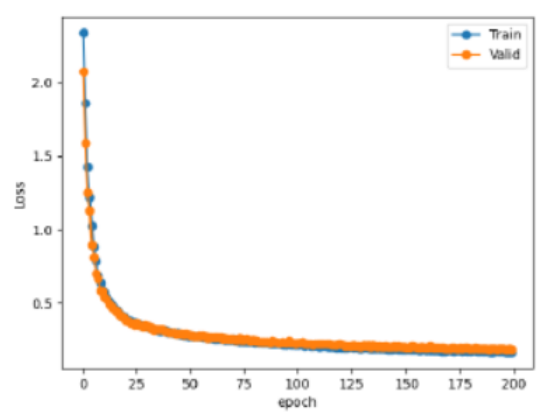
Deep Learning-Based Hand Pose Estimation: A Novel ResUnet Framework for Occlusion Handling in AR/VR Applications

Roumaissa BEKIRI, Mohamed Chaouki BABAHENINI
LESIA Laboratory, Mohamed Khider University, Biskra, Algeria.

GANerated dataset
In particular, we have used the mean square error (MSE) as a pivotal metric to assess errors during the training and validation stages on the GANerated dataset.
Our comprehensive training approach involved 200 epochs, 100 batches, and 64 batches per epoch. We used a stochastic gradient descent optimizer with a momentum value 0.9 and a learning rate of 0.005. In addition, the weights in the ResNet network were kept frozen throughout the training process. These experimental settings resulted in notable improvements in loss, as depicted in Figure 4(b).
During the data augmentation process, we provide the same hyperparameters as those employed during the training. Our investigations have revealed intriguing disparities in the impact of augmentation on 2D and 3D training errors when applied to the GANerated and SynthHands datasets, which we have noted in Fig.5



(a) 2D and 3D PCK metrics under the perspective threshold of our architecture.



(b) Mean Square Error(MSE) applied during training and validation test data.

Fig. 4: Quantitative Evaluation of our proposed approach on GANerated dataset.

Deep Learning-Based Hand Pose Estimation: A Novel ResUnet Framework for Occlusion Handling in AR/VR Applications

Roumaissa BEKIRI, Mohamed Chaouki BABAHENINI
LESIA Laboratory, Mohamed Khider University, Biskra, Algeria.

As indicated in Fig.6, our approach has achieved a remarkable AUC value of 0.945, demonstrating a significant advancement compared to the work of Mueller et al.[6]

		Metrics			
		2D mEPE	3D mEPE	2D AUC	3D AUC
Before Augmentation	GANerated dataset	0.063	0.1718	0.93	0.823
	SynthHands datasets	0.034	0.134	0.960	0.860
After Augmentation	GANerated dataset	0.058	0.165	0.94	0.842
	SynthHands datasets	0.041	0.1842	0.953	0.811

Fig. 5: The effect of data augmentation on the GANerated and SynthHands datasets using evaluation metrics: 2D and 3D mean End Point Error(mEPE) and 2D and 3D AUC.

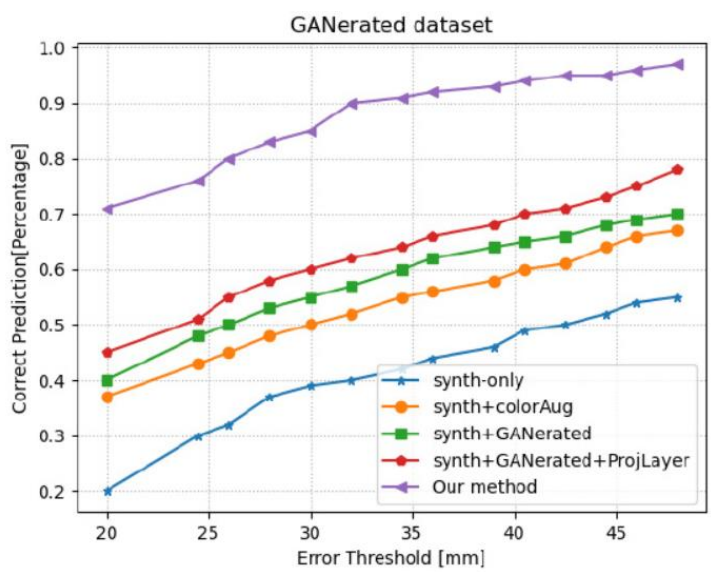


Fig. 6: 3D PCK on GANerated datasets. Comparison with the work of Mueller et al.[6] in a different manner.

Roumaissa BEKIRI, Mohamed Chaouki BABAHENINI

LESIA Laboratory, Mohamed Khider University, Biskra, Algeria.

In Fig.7, the average mEPE of approximately 4.94 mm demonstrates the successful distribution of errors for each joint during training.

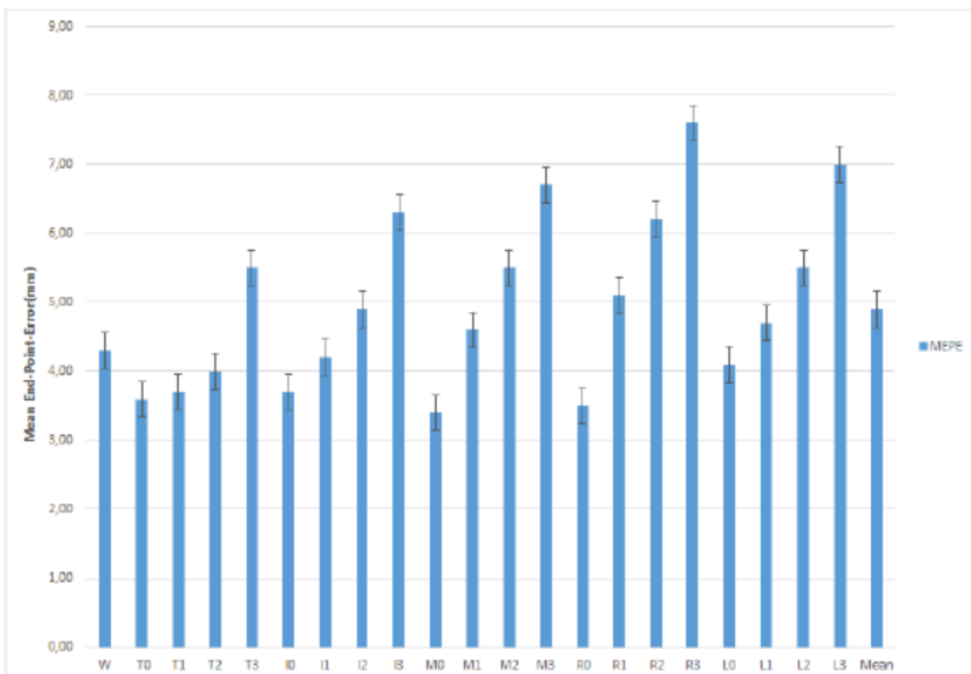


Fig. 7: The Mean End-Point-Error of each joint of a full hand on GANerated dataset [7].

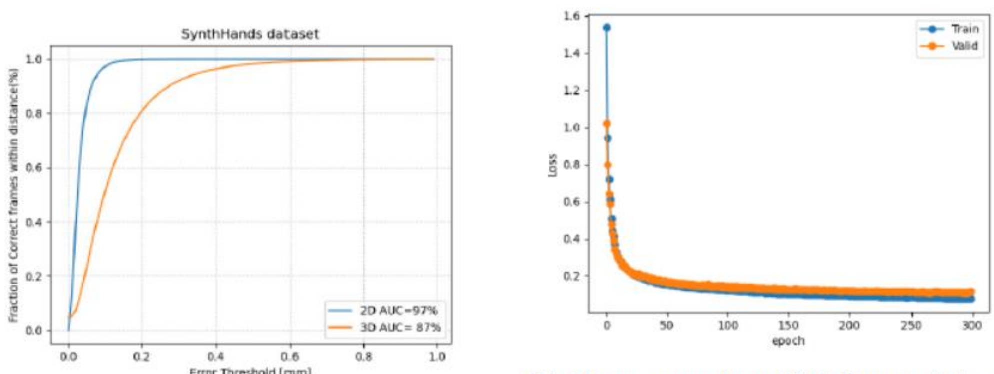
SynthHands datasets

Furthermore, regarding loss, Figure 8.(b) illustrates the Mean Square Error(MSE) throughout the training and validation phases. The plot indicates an optimized value as the number of training epochs increase, reflecting the learning process's effectiveness.

Deep Learning-Based Hand Pose Estimation: A Novel ResUnet Framework for Occlusion Handling in AR/VR Applications

Roumaissa BEKIRI, Mohamed Chaouki BABAHENINI
LESIA Laboratory, Mohamed Khider University, Biskra, Algeria.

We used the 3D Mean End-Point Error to compare our model with the work of Li et al.[8] across various testing scenarios. As shown in Fig.9, our model consistently outperforms the competition, with a mean Euclidean distance error of 6.33 mm.



(a) The successful pose estimation frames in 2D and 3D under different error thresholds.

(b) Mean square Error(MSE) on training and validation dataset with appropriate hyperparameters.

Fig. 8: Quantitative evaluation of our proposal on SynthHands datasets.

Stereo Dataset

As shown in Fig.10, we compared the recent work. Li et al. [8] and Iqbal et al. [9] achieved mEPE values of about 3.57 mm and 3.54 mm, respectively. In comparison, our model achieved a lower mEPE of 2.96mm, highlighting its enhanced accuracy in predicting hand keypoints.

However, we conducted a comparative analysis with recently proposed methods, utilizing the 3D Percentage of Correct Keypoints (PCK), as depicted in Fig.11. Our analysis revealed that our model achieved a remarkable AUC within the range of 20-50 mm, approximately 0.999, surpassing the performance of all recent approaches.

Deep Learning-Based Hand Pose Estimation: A Novel ResUnet Framework for Occlusion Handling in AR/VR Applications

Roumaissa BEKIRI, Mohamed Chaouki BABAHENINI
LESIA Laboratory, Mohamed Khider University, Biskra, Algeria.

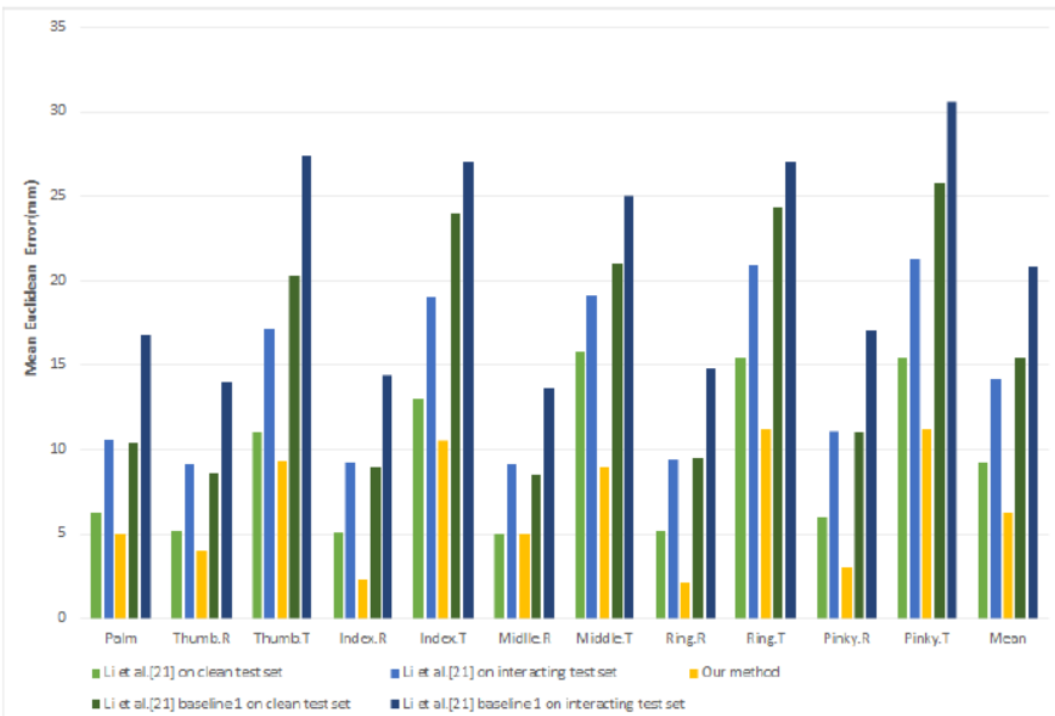


Fig. 9: Mean End-point-Error of different joints of full hand on SynthHands Dataset.

Method	2D mEPE(mm)(in pixel) ↓	AUC of 2D PCK_{0-50}
Li et al. [8]	13.58	-
Zimmermann et al. [10]	5.522	0.817
kourbane et al.	6.21	0.796
Gao et al.	5.059	0.831
Yuan et al.	5.801	-
Ours	**2.96**	**0.942**

Fig. 10: 2D comparison with the existing methods on STB dataset utilizing Mean EPE[mm](pixels) and 2D AUC. "↓" indicates that lower is good, "↑": higher is better

Deep Learning-Based Hand Pose Estimation: A Novel ResUnet Framework for Occlusion Handling in AR/VR Applications

Roumaissa BEKIRI, Mohamed Chaouki BABAHENINI
LESIA Laboratory, Mohamed Khider University, Biskra, Algeria.

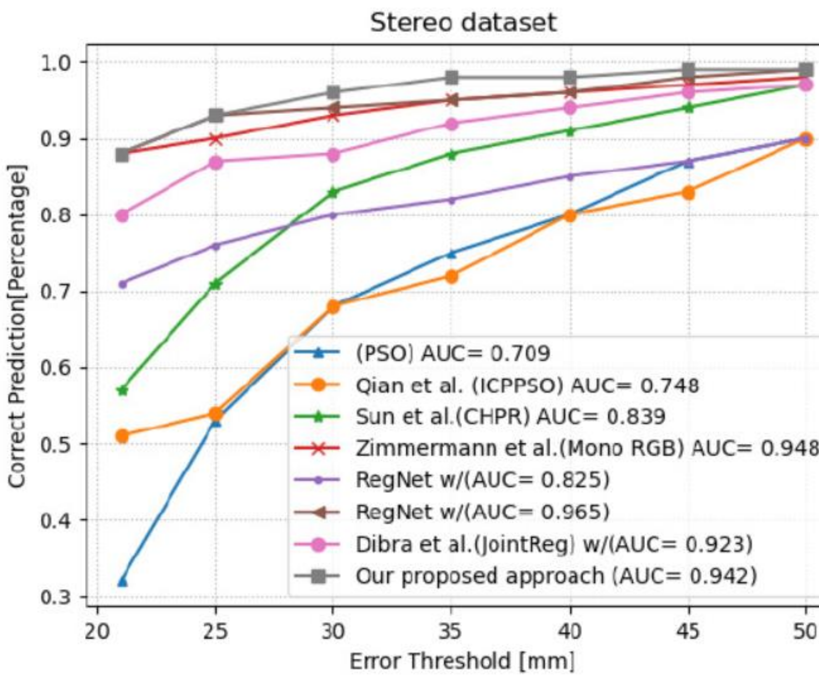


Fig. 11: Comparison with the state-of-the-art work[7, 10–13]on the STB dataset utilizing 3D PCK. The X-axis is the threshold values, and Y-axis is the 3D PCK over the perspective threshold. The Area Under the Curve in this curve is AUC(1-100)[mm].

3.2 Qualitative evaluation

In addition to the quantitative assessments, we conducted a qualitative evaluation on three available datasets: GANerated, SynthHands, and STB dataset.

Deep Learning-Based Hand Pose Estimation: A Novel ResUnet Framework for Occlusion Handling in AR/VR Applications

Roumaissa BEKIRI, Mohamed Chaouki BABAHENINI
LESIA Laboratory, Mohamed Khider University, Biskra, Algeria.

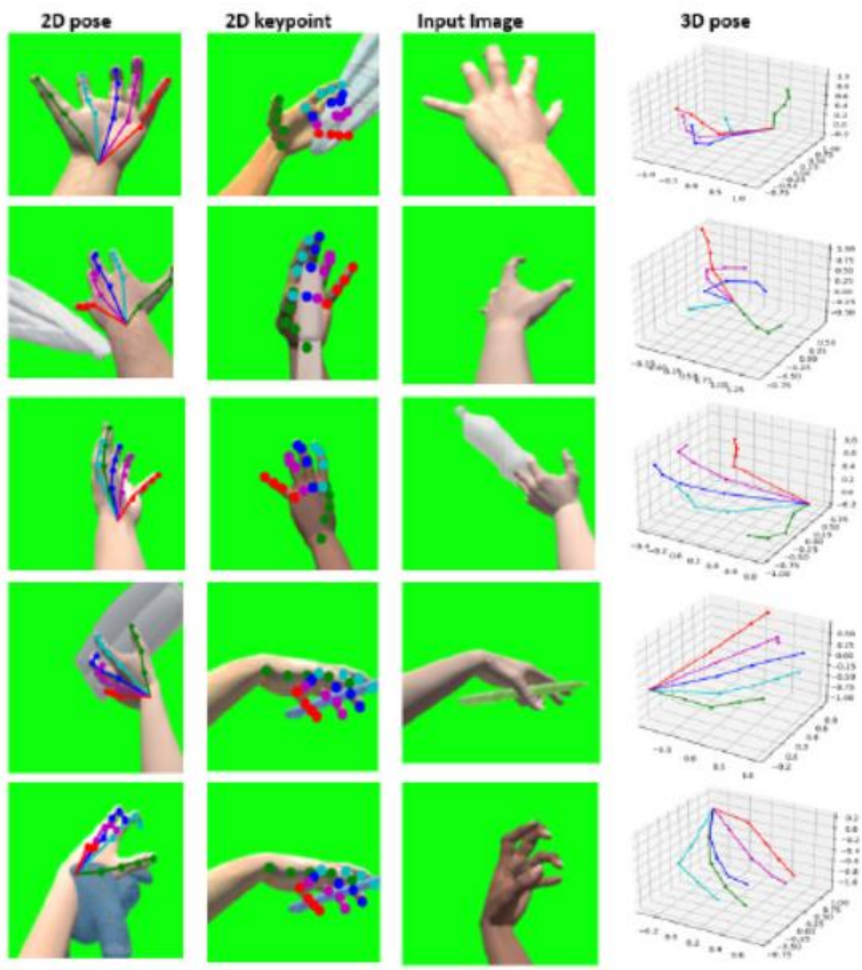


Fig. 12: Qualitative results on SynthHANDS dataset [6]. The first and second rows are images estimated from the proposed approach with 2D keypoints and a 2D skeleton. The third one represents some images from a dataset, and the following row shows the estimated 3D joints of the hand.

Deep Learning-Based Hand Pose Estimation: A Novel ResUnet Framework for Occlusion Handling in AR/VR Applications

Roumaissa BEKIRI, Mohamed Chaouki BABAHENINI
LESIA Laboratory, Mohamed Khider University, Biskra, Algeria.

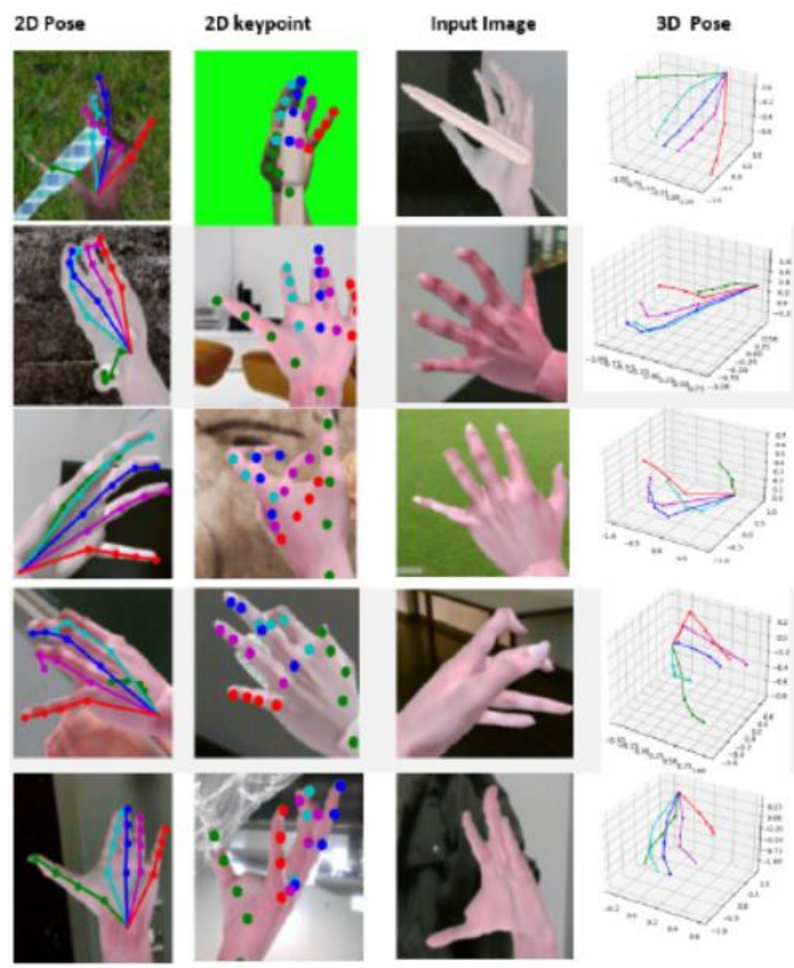


Fig. 13: Qualitative results on GANerated dataset [7]. The initial and the second columns are images estimated from the proposed approach with 2D keypoints and 2D skeleton. The third one represents some hand images, and the following column shows the estimated 3D joints of the hand.

Deep Learning-Based Hand Pose Estimation: A Novel ResUnet Framework for Occlusion Handling in AR/VR Applications

Roumaissa BEKIRI, Mohamed Chaouki BABAHENINI
LESIA Laboratory, Mohamed Khider University, Biskra, Algeria.

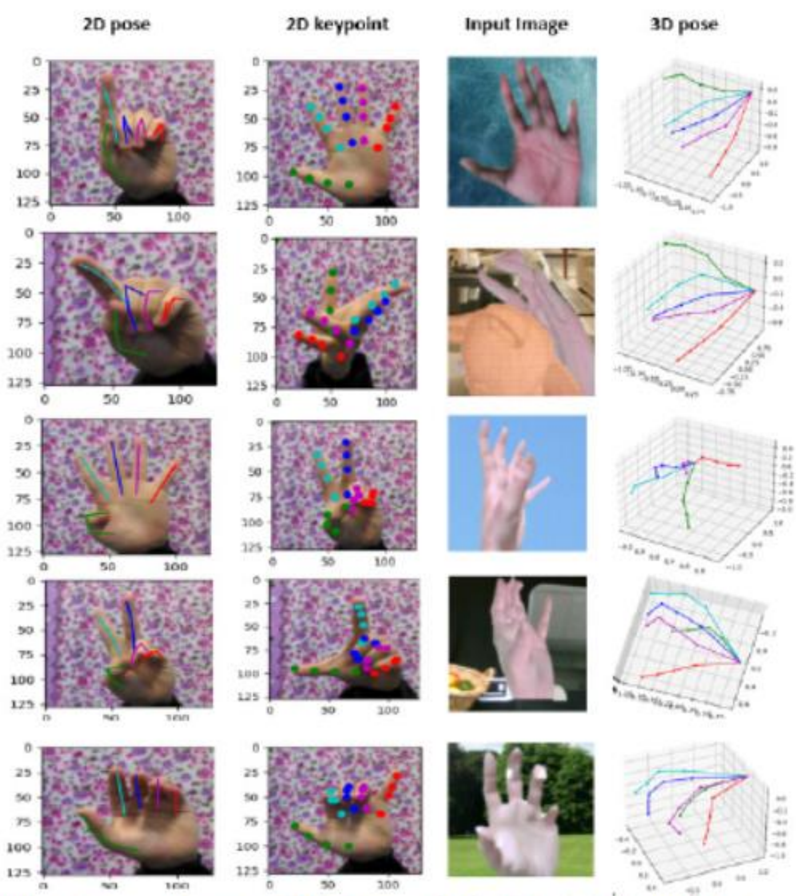


Fig. 14: Qualitative results on STB dataset[14]. The initial and the second columns are some images estimated from our proposed approach with 2D keypoints and a 2D skeleton. The third one represents the hand image, and the following column shows the estimated 3D joints of the hand.

Visual Computing Magazine

Deep Learning-Based Hand Pose Estimation: A Novel ResUnet Framework for Occlusion Handling in AR/VR Applications

Roumaissa BEKIRI, Mohamed Chaouki BABAHENINI
LESIA Laboratory, Mohamed Khider University, Biskra, Algeria.

4 Conclusion and future work

Hand pose estimation is a critical challenge in computer vision, particularly in augmented and virtual reality applications. This paper presents a new deep learning framework, ResUnet, which combines the strengths of the ResNet-34 and U-Net architectures to efficiently estimate 2D and 3D hand poses from single RGB images. We carried out extensive experiments on benchmark datasets, such as GANerated, SynthHands, and STB, to validate the effectiveness of the proposed approach.

Quantitative results show superior performance, outperforming existing methods, while qualitative assessments confirm the model's ability to accurately predict key hand points and reconstruct 3D poses under challenging conditions.

Future work could explore real-time optimization for deployment on peripheral devices and generalization to more diverse manual interactions, such as multi-hand scenarios or dynamic gestures.

References

[1] Piumsomboon, T., Clark, A., Billinghurst, M., Cockburn, A.: User-defined gestures for augmented reality. In: IFIP Conference on Human-Computer Interaction, pp. 282–299 (2013). Springer
[2] Yin, F., Chai, X., Chen, X.: Iterative reference driven metric learning for signer independent isolated sign language recognition. In: European Conference on Computer Vision, pp. 434–450 (2016). Springer
[3] Markussen, A., Jakobsen, M.R., Hornbæk, K.: Vulture: a mid-air word-gesture keyboard. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 1073–1082 (2014)
[4] Jang, Y., Noh, S.-T., Chang, H.J., Kim, T.-K., Woo, W.: 3d finger cape: Clicking action and position estimation under self-occlusions in egocentric viewpoint. IEEE Transactions on Visualization and Computer Graphics 21(4), 501–510 (2015)
[5] Le, V.-H., Nguyen, H.-C.: A survey on 3d hand skeleton and pose estimation by convolutional neural network. Adv Sci Technol Eng Syst J 5, 144–159 (2020)

Deep Learning-Based Hand Pose Estimation: A Novel ResUnet Framework for Occlusion Handling in AR/VR Applications

Roumaissa BEKIRI, Mohamed Chaouki BABAHENINI
LESIA Laboratory, Mohamed Khider University, Biskra, Algeria.

[6] Mueller, F., Mehta, D., Sotnychenko, O., Sridhar, S., Casas, D., Theobalt, C.: Real-time hand tracking under occlusion from an egocentric rgb-d sensor. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 1154–1163 (2017)
[7] Mueller, F., Bernard, F., Sotnychenko, O., Mehta, D., Sridhar, S., Casas, D., Theobalt, C.: Ganerated hands for real-time 3d hand tracking from monocular rgb. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 49–59 (2018)
[8] Li, M., Wang, J., Sang, N.: Latent distribution-based 3d hand pose estimation from monocular rgb images. IEEE Transactions on Circuits and Systems for Video Technology 31(12), 4883–4894 (2021)
[9] Iqbal, U., Molchanov, P., Gall, T.B.J., Kautz, J.: Hand pose estimation via latent 2.5d heatmap regression. In: Proceedings of the European Conference on Computer Vision (ECCV), pp. 118–134 (2018)
[10] Zimmermann, C., Brox, T.: Learning to estimate 3d hand pose from single rgb images. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 4903–4911 (2017)
[11] Chen, L., Lin, S.-Y., Xie, Y., Tang, H., Xue, Y., Xie, X., Lin, Y.-Y., Fan, W.: Generating realistic training images based on tonality-alignment generative adversarial networks for hand pose estimation. arXiv preprint arXiv:1811.09916 (2018)
[12] Qian, C., Sun, X., Wei, Y., Tang, X., Sun, J.: Realtime and robust hand tracking from depth. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1106–1113 (2014)
[13] Dibra, E., Melchior, S., Balkis, A., Wolf, T., Oztireli, C., Gross, M.: Monocular rgb hand pose inference from unsupervised refinable nets. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, pp. 1075–1085 (2018)
[14] Zhang, Y., Xu, C., Cheng, L.: Learning to search on manifolds for 3d pose estimation of articulated objects. arXiv preprint arXiv:1612.00596 (2016)

Visual Computing Magazine

Machine and Deep learning Techniques for Histological and Endoscopic Images

Abdelouahab MOUSSAOUI, Computer Science Department
Faculty of Sciences, Ferhat Abbas University - Sétif 1, Algeria

Abstract

Colorectal cancer (CRC) or we can call it colon cancer, it is known among the most common type of cancer in the world, according to the latest statistics, CRC is more common in men compared to women. For early detection, endoscopic imaging is used to determine the presence of colonic polyps, and for accurate confirmation and diagnosis of CRC, a histological examination is used to study and analyze the colonic tissues, with a purpose to classify the abnormal growth as malignant tumor and the normal growth (colonic polyps) as benign tumor. In this paper, we develop both Machine learning and Deep learning models in supervised frameworks Due to the existence of labeled dataset with the classes of colonic tissues that can classify medical images (histological colonic tissues) to improve the diagnosis and treatment of colorectal cancer in order to be able to make difference between healthy colonic tissue and cells affected by cancer.

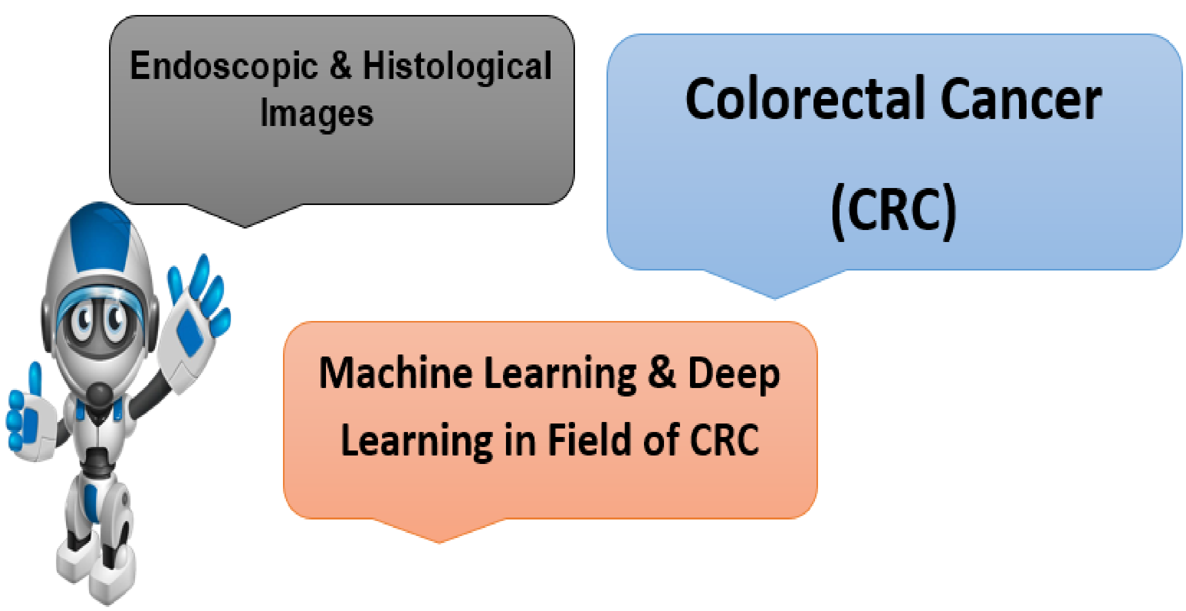
Index Terms—Colorectal cancer(CRC); Colorectal polyps; Histological images; Machine learning; Deep learning

Machine and Deep learning Techniques for Histological and Endoscopic Images

Abdelouahab MOUSSAOUI, Computer Science Department
Faculty of Sciences, Ferhat Abbas University - Sétif 1, Algeria

I. INTRODUCTION

CRC is one of the most dangerous types of cancer that leads to death, it directly targets the colon, rectum, or both. It is the 3rd most commonly diagnosed cancer worldwide. In our study, we will illustrate our contribution studies using many techniques of artificial intelligence to improve the diagnosis and treatment of colorectal cancer. We will investigate histological and endoscopic images to develop predictive models to distinguishing between colonic tissues to identify patients at risk of developing colorectal cancer, as well as to predict patient outcomes. Ultimately, we hope to contribute to the development of more accurate and efficient methods for colorectal cancer diagnosis and treatment.



Machine and Deep learning Techniques for Histological and Endoscopic Images

Abdelouahab MOUSSAOUI, Computer Science Department
Faculty of Sciences, Ferhat Abbas University - Sétif 1, Algeria

II. OBJECTIVES AND MOTIVATIONS

In our work, we proposed supervised models for binary classification of colonic tissues; in another study, we proposed supervised models for semantic segmentation models of endoscopic images. The challenges that we face are that histological image resources are very rare and the private datasets, also the lack of existence of masks associated with histological images.

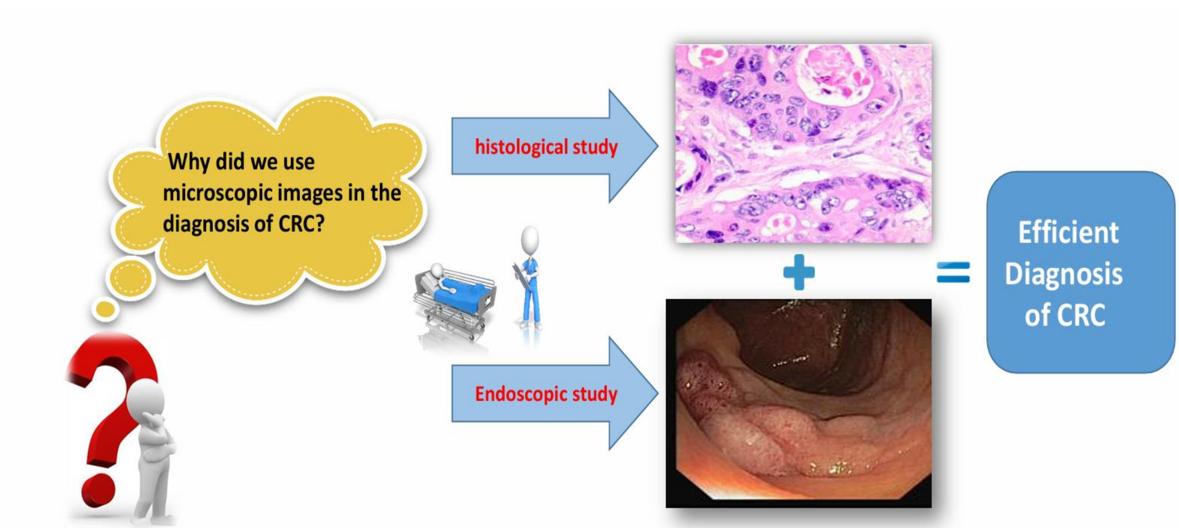


Fig. 1. Workflow of our contribution

Problem Formulation

We formulate our problem that is classified among the most common medical problems related to colon cancer, which were included in the framework of artificial intelligence problems to integrate the medical field with AI to obtain more accurate results regarding the Classification of colonic tissues in Histological images, and Segmentation of colonic polyps in Colonoscopic images.

Machine and Deep learning Techniques for Histological and Endoscopic Images

Abdelouahab MOUSSAOUI, Computer Science Department
Faculty of Sciences, Ferhat Abbas University - Sétif 1, Algeria

we tried to create machine and deep learning models to solve those problems that we talked about just in the problem formulation, so we create models concerning segmentation task for endoscopic images and models for classification task for histological images:

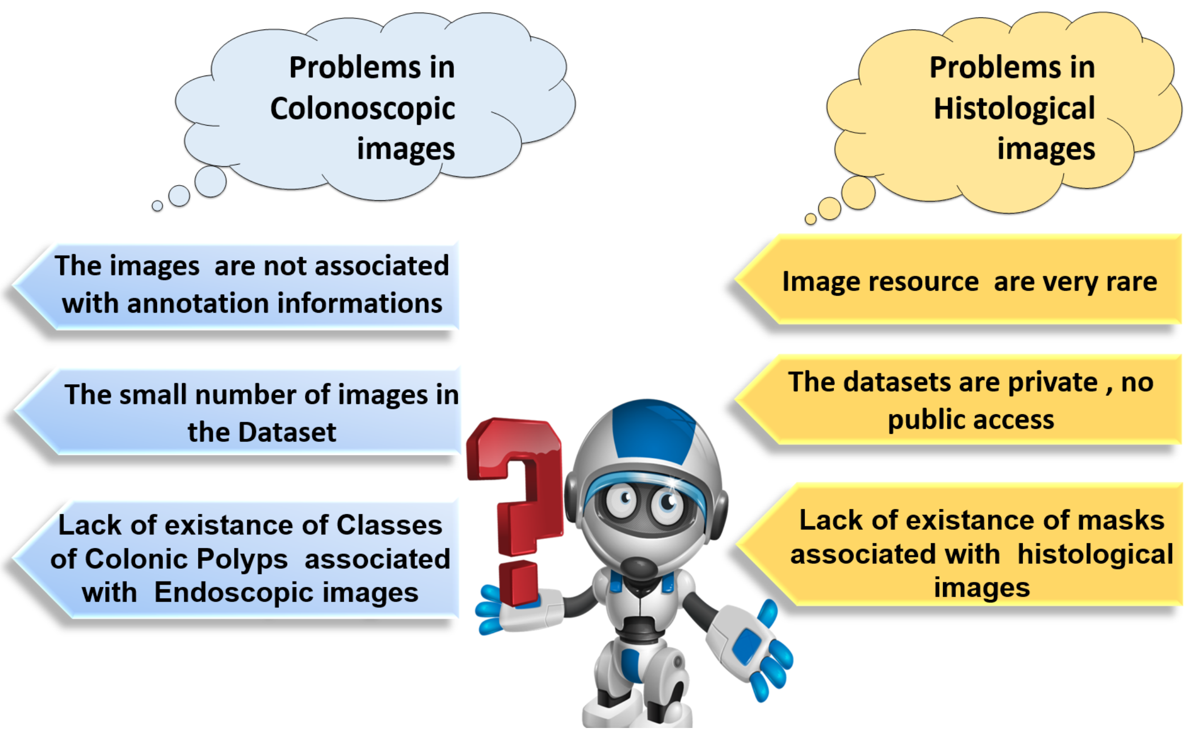


Fig. 2. Problem Formulation

III. MATERIALS AND METHODS

A. Data Collections

1. Dataset of histological images: We used LC25000 Dataset of histological color images, each classes contain 5000 images in 768 x768 pixels with jpeg format. The classes of colon tissues: colon-aca with 5000 images that represent colon adeno-carcinomas and colon-n with 5000 images of benign colonic tissues.

Machine and Deep learning Techniques for Histological and Endoscopic Images

Abdelouahab MOUSSAOUI, Computer Science Department
Faculty of Sciences, Ferhat Abbas University - Sétif 1, Algeria

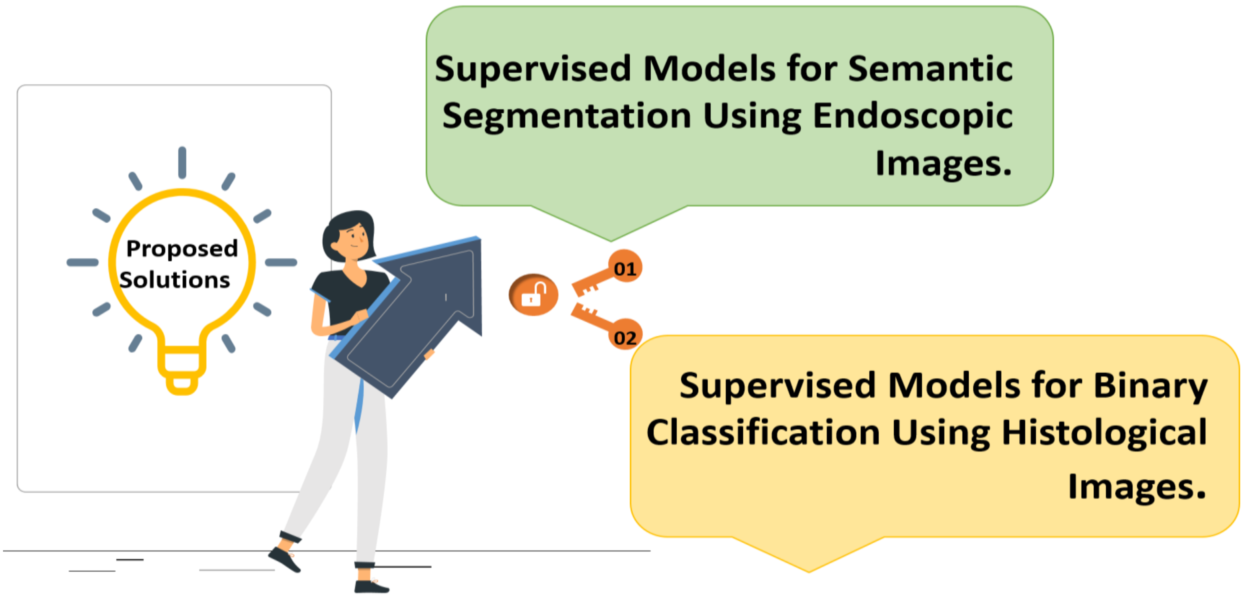


Fig. 3. Proposed Solutions

2. Dataset of Endoscopic images: The dataset of endoscopic images is a database of frames extracted from colonoscopy videos from the Cancer Hospital in Setif Province(CAC). These frames contain several examples of polyps. In addition to the frames, we provide the ground truth for the polyps. This ground truth consists of a mask corresponding to the region covered by the polyp in the image. In this image below, First row shows original images whereas second row shows corresponding ground truth. The medical data images were anonymized, to insure the medical secret and protect the confidentiality.

B. Dataset Splitting

For the first split of our histological dataset, we split it into 0.8 for training-set which represent 8000 images for both classes, and 0.2 for testing-set which represent 2000 images. For the second splitting, from the training-set we make 0.15for validation-set which represent 1200 images, and the rest of dataset still for training which represent 6800 images.

Visual Computing Magazine

Machine and Deep learning Techniques for Histological and Endoscopic Images

Abdelouahab MOUSSAOUI, Computer Science Department
Faculty of Sciences, Ferhat Abbas University - Sétif 1, Algeria

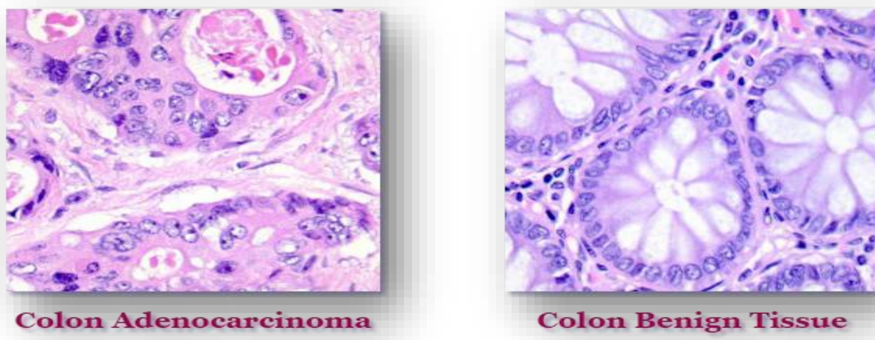


Fig. 4. Dataset of histological images

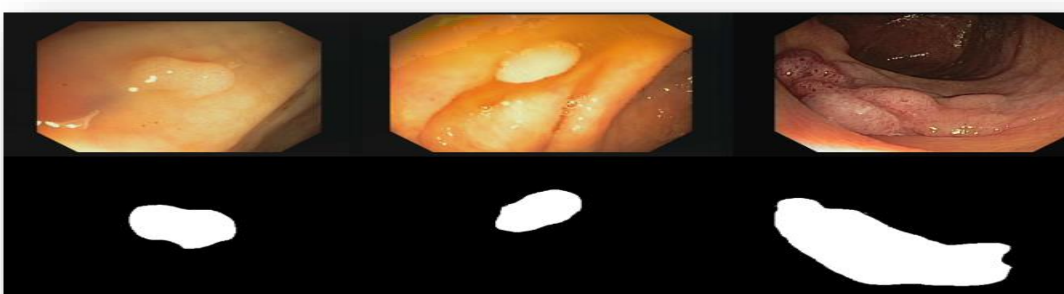


Fig. 4. Dataset of Endoscopic images

Visual Computing Magazine

Machine and Deep learning Techniques for Histological and Endoscopic Images

Abdelouahab MOUSSAOUI, Computer Science Department
Faculty of Sciences, Ferhat Abbas University - Sétif 1, Algeria

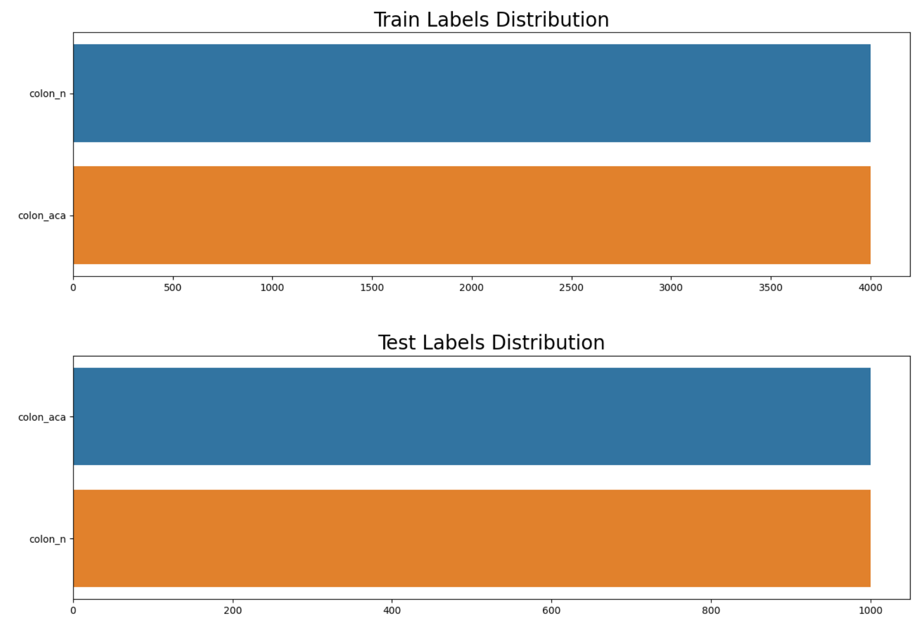


Fig. 5. dataset splitting 1

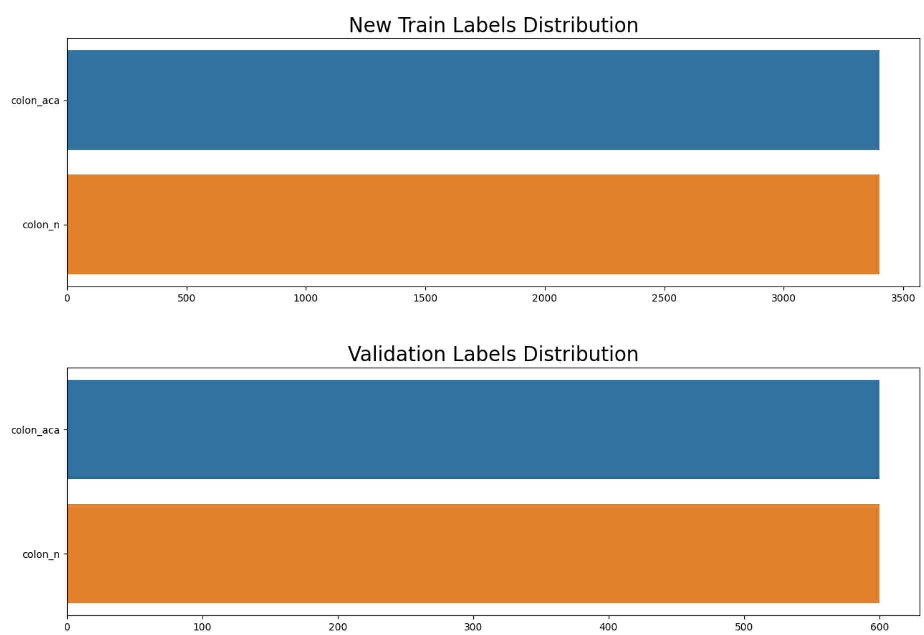


Fig. 6. dataset splitting 2

Machine and Deep learning Techniques for Histological and Endoscopic Images

Abdelouahab MOUSSAOUI, Computer Science Department
Faculty of Sciences, Ferhat Abbas University - Sétif 1, Algeria

C. Pre-processing and Data Exploration

After splitting our dataset into 3 parts: train-set, validation-set and test-set. we generate labels for the images based on their file paths, creates a Pandas Data Frame containing the image paths and labels, encodes the labels, by adding
a new column called" label-encoded". The label encoding is based on the 'label' column; if the label is" colon-aca", the encoding is 0, and if the label is anything else, the encoding is 1 in our case it is for "colon-n". Finally, the shuffled Data Frame is reset and returned. Then we Create Input Data Pipeline for a TensorFlow model by extracts the image paths and labels from the Data Frame and creates a dataset. It then applies data augmentation and the image loading function to the dataset inputs, based on the augment flag. Next, the function applies shuffling, batching, caching,
and prefetching to the dataset, based on the arguments passed to the function. Finally, the function returns the dataset, which can be used to train a TensorFlow model.

	image_path	label	label_encoded
0	/content/gdrive/MyDrive/Colab Notebooks/train/...	colon_n	1
1	/content/gdrive/MyDrive/Colab Notebooks/train/...	colon_n	1
2	/content/gdrive/MyDrive/Colab Notebooks/train/...	colon_n	1
3	/content/gdrive/MyDrive/Colab Notebooks/train/...	colon_n	1
4	/content/gdrive/MyDrive/Colab Notebooks/train/...	colon_aca	0

Fig. 7. Label Generation and Encoding

Visual Computing Magazine

Machine and Deep learning Techniques for Histological and Endoscopic Images

Abdelouahab MOUSSAOUI, Computer Science Department
Faculty of Sciences, Ferhat Abbas University - Sétif 1, Algeria

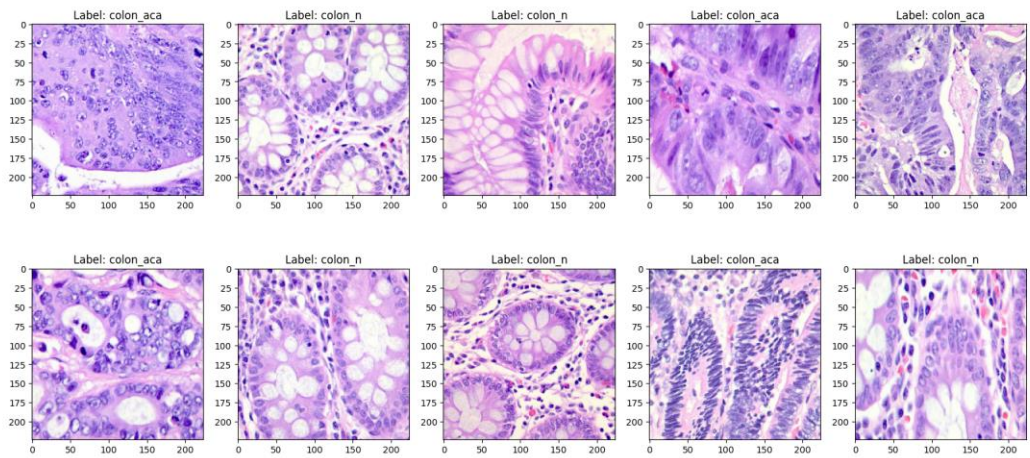


Fig. 8. Data Exploration

CONTRIBUTION STUDIES

Here we illustrate the models that we proposed to make the semantic segmentation of colonic polyps.

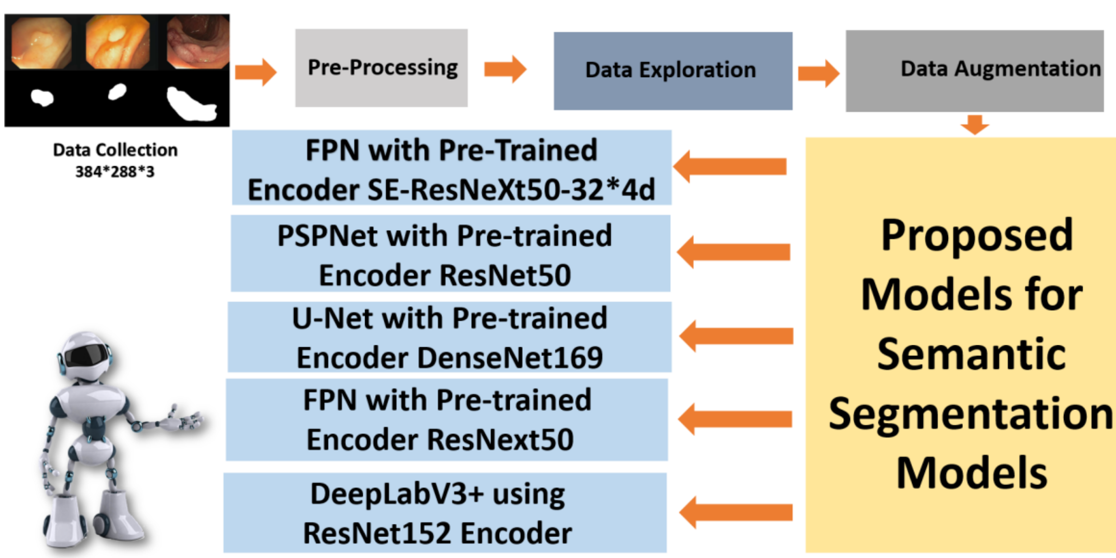


Fig. 9. Proposed Models of Colonoscopic images

Machine and Deep learning Techniques for Histological and Endoscopic Images

Abdelouahab MOUSSAOUI, Computer Science Department
Faculty of Sciences, Ferhat Abbas University - Sétif 1, Algeria

In other hand, we represent our proposed architectures for binary classification of colonic tissues.

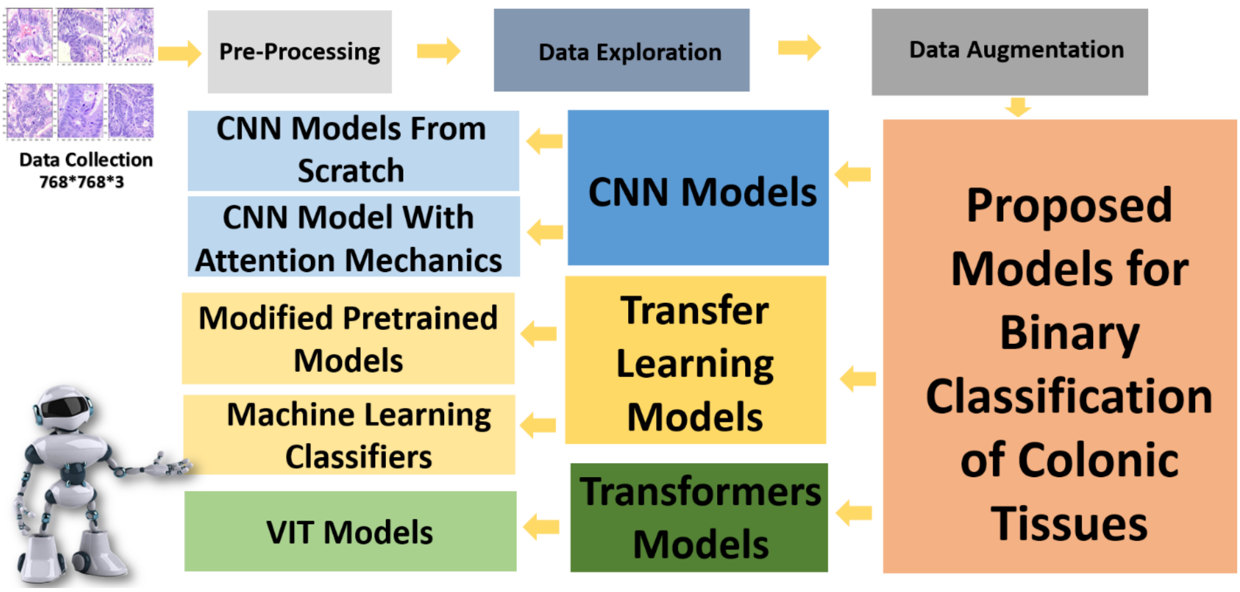


Fig. 10. Proposed Models of Histological images

Here, we detail the mechanism for creating encoding labels for colonoscopic masks.

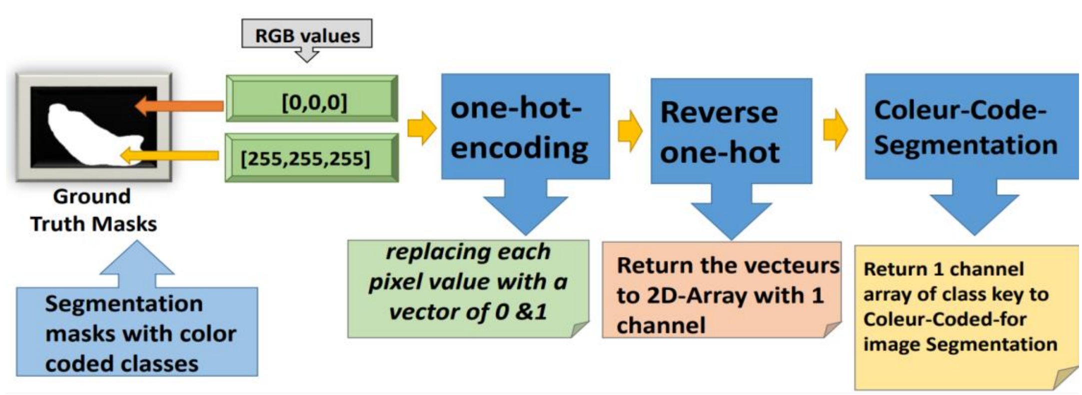


Fig. 11. Process of Encoding Labels of Colonoscopic masks

Machine and Deep learning Techniques for Histological and Endoscopic Images

Abdelouahab MOUSSAOUI, Computer Science Department
Faculty of Sciences, Ferhat Abbas University - Sétif 1, Algeria

In our case the classification is binary which means our output is in binary value either 0 or 1, which represent in our model the type of tissue: colon-aca or colon-n.
Here we visualize the workflow of The Process of CNN Models:

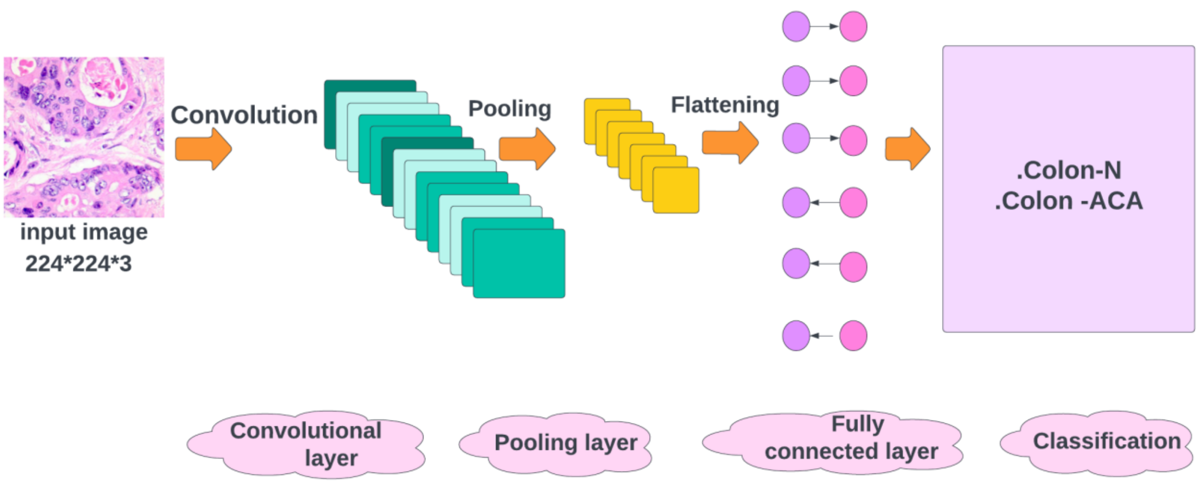


Fig. 12. workflow of CNN models

Also, we mentioned in details the mechanism of transfer learning, we used MPMs for image classification task by loading the pretrained models, and replace the final classification layer with new layer to make specific task, this new layer is trained on our dataset, don't miss that the weight of pretrained layers are frozen or finetuned, it is depending to the size of our dataset. Finally, we found that the main purpose of using MPMs is to augment the performances of the model on specific task comparing to training from the scratch.

Machine and Deep learning Techniques for Histological and Endoscopic Images

Abdelouahab MOUSSAOUI, Computer Science Department
Faculty of Sciences, Ferhat Abbas University - Sétif 1, Algeria

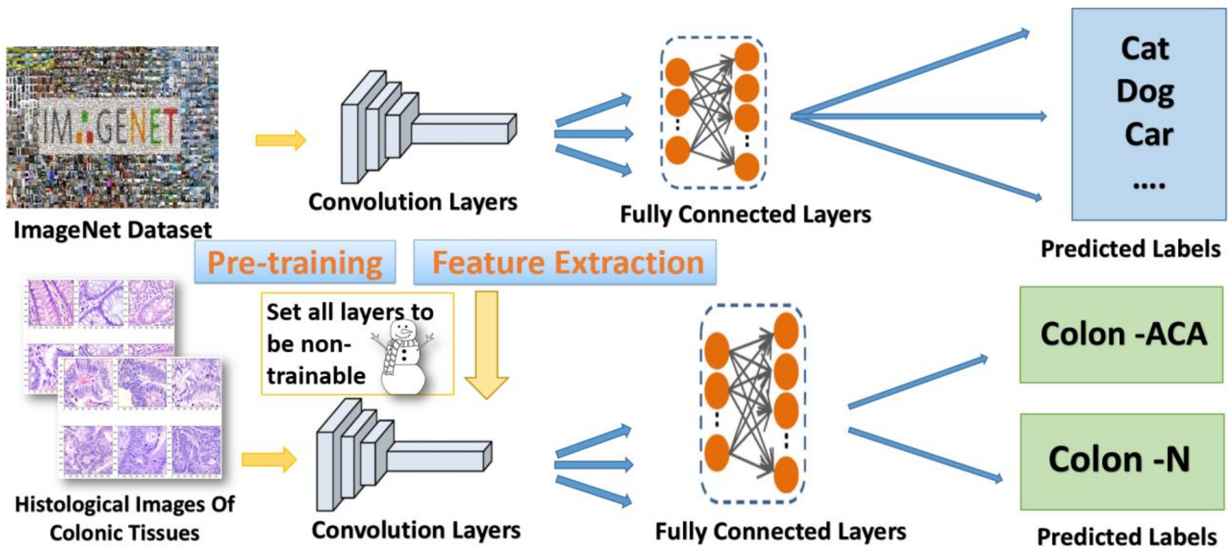


Fig. 13. workflow of transfer learning Models

we passed the input images of colonic tissues to the Modified pretrained Resnet50, to make the feature extraction step, then we pass the extracted features to the classification layer which represent machine learning classifiers such as KNN, SVC and Random forest, and finally we display our output classes, colon-aca with 0 and colon-n with 1.

Machine and Deep learning Techniques for Histological and Endoscopic Images

Abdelouahab MOUSSAOUI, Computer Science Department
Faculty of Sciences, Ferhat Abbas University - Sétif 1, Algeria

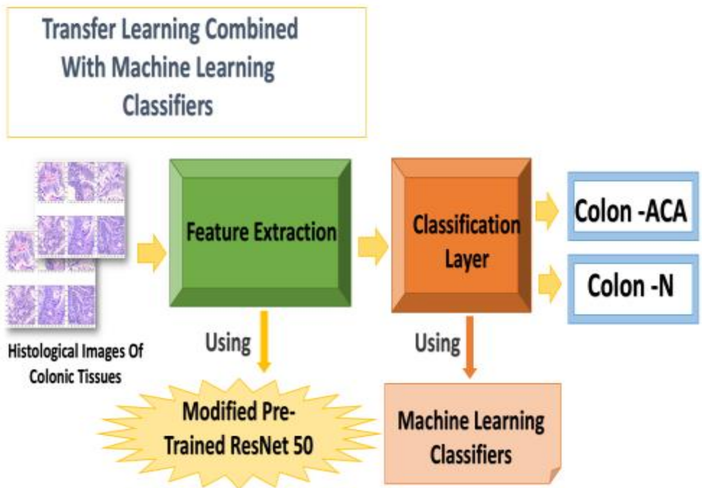


Fig. 14. workflow of Machine Learning classifiers

VIT is a powerful architecture for image classification tasks because it can handle images of arbitrary sizes, by using multi-head self-attention, capture long-range dependencies between different regions of an image, and learn feature representations using self-attention mechanisms, we used VIT B16 and VIT B32.

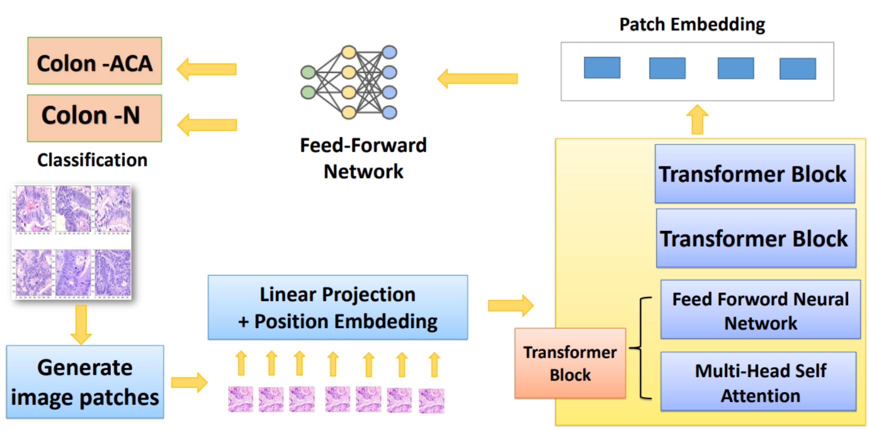


Fig. 15. Workflow of VIT models

Machine and Deep learning Techniques for Histological and Endoscopic Images

Abdelouahab MOUSSAOUI, Computer Science Department
Faculty of Sciences, Ferhat Abbas University - Sétif 1, Algeria

VI. EXPERIMENTAL STUDIES AND RESULTS OF CLASSIFICATION MODELS

The CNN with Convolutional Channel Attention (CBAM) model is performing very well on the dataset. The loss is very low just 0.00698 and the accuracy with 0.997, precision with 0.998, recall with 0.996. This means that our model is able to classify the images with high accuracy, and is able to correctly identify both positive and negative examples. We used Gradient-weighted Class Activation Mapping, to highlights the most important regions of an image, that led to a particular classification decision by visualizing the class activation maps of the CNN which represent the most important regions, in producing the final classification decision

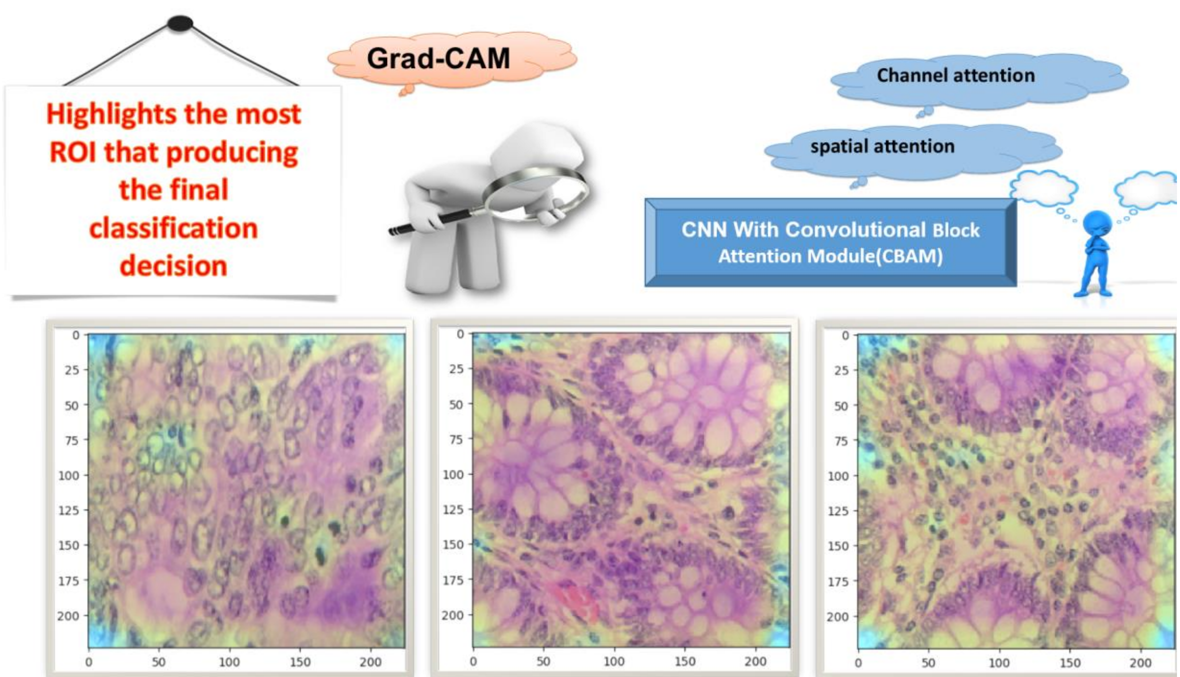


Fig. 16. Grad Cam Visualization of CNN Model with Attention

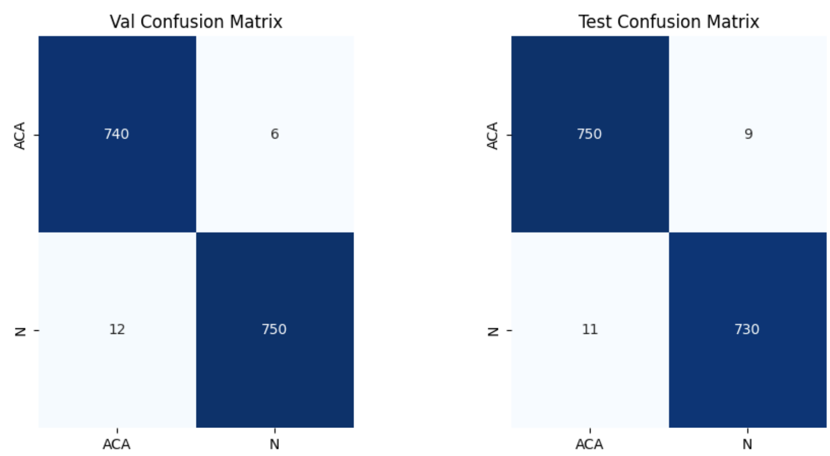


Fig.17. Result of Modified VGG-19 Pre-trained Model

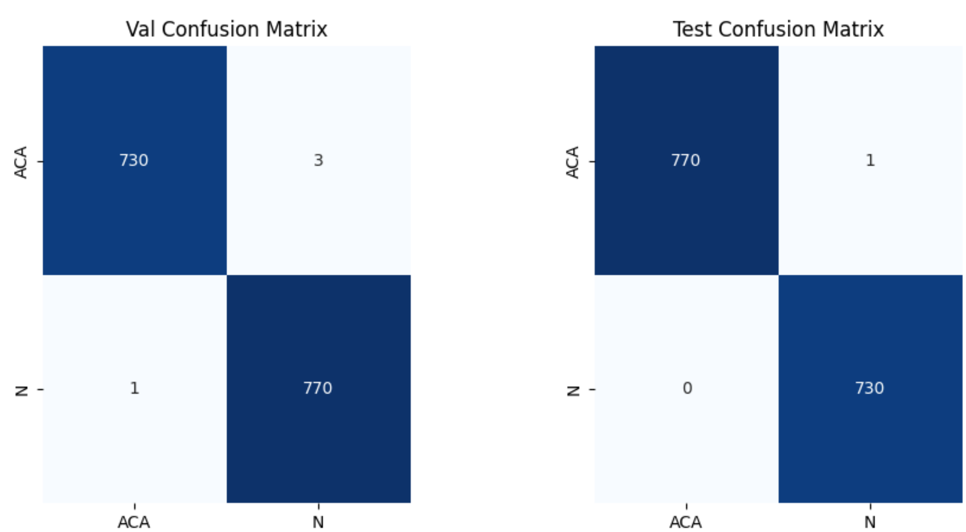


Fig.18. Results of Modified ResNet50 Combined with SVM Classifier

Machine and Deep learning Techniques for Histological and Endoscopic Images

Abdelouahab MOUSSAOUI, Computer Science Department
Faculty of Sciences, Ferhat Abbas University - Sétif 1, Algeria

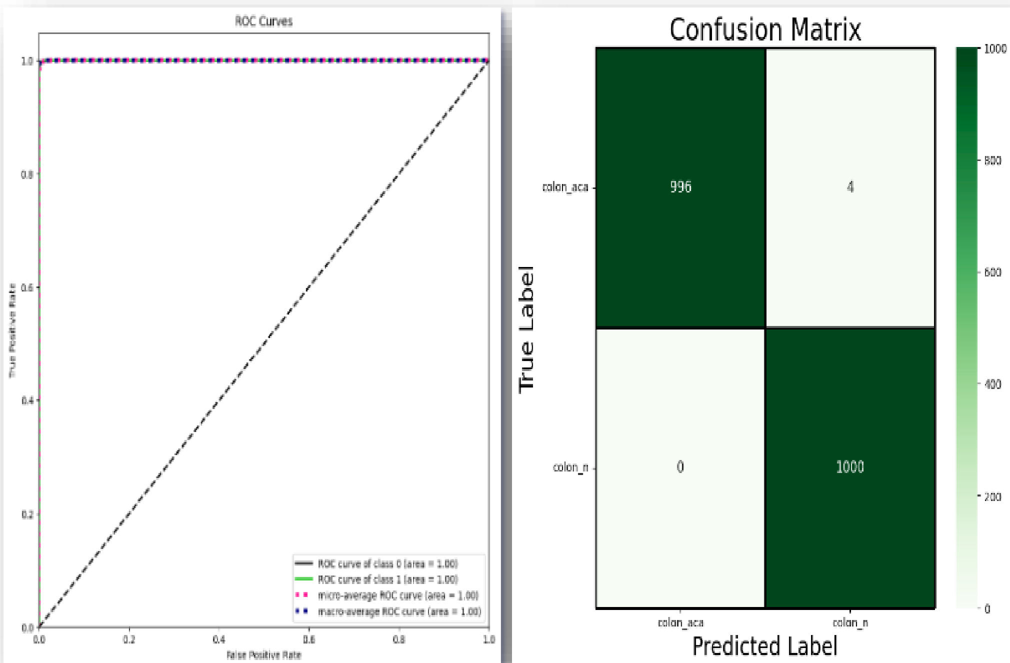


Fig.19. Result of VIT B16 model

EXPERIMENTAL STUDIES AND RESULTS OF SEGMENTATION MODELS

We display some results of our proposed models to make the semantic segmentation using endoscopic images

Machine and Deep learning Techniques for Histological and Endoscopic Images

Abdelouahab MOUSSAOUI, Computer Science Department
Faculty of Sciences, Ferhat Abbas University - Sétif 1, Algeria

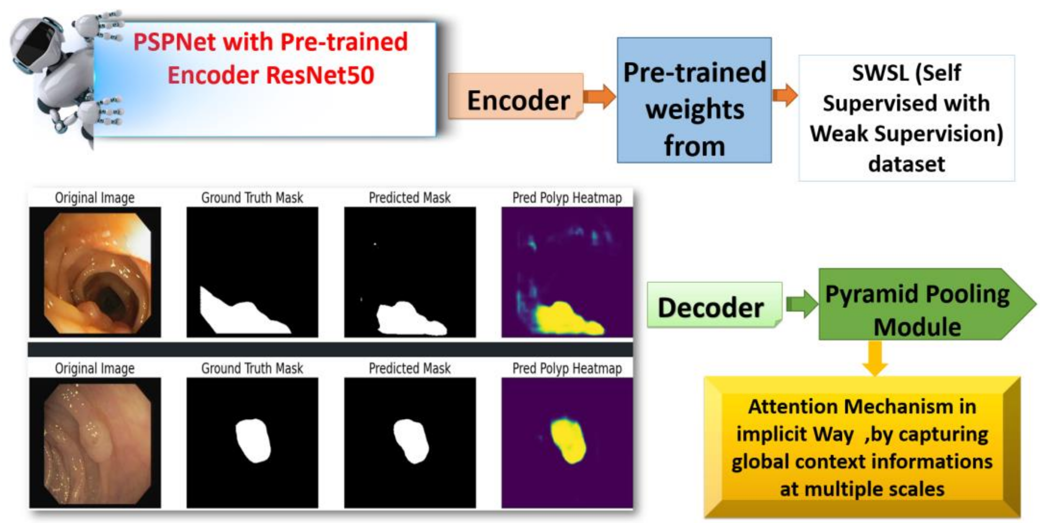


Fig.20. PSPNet with Pre-trained Encoder ResNet50

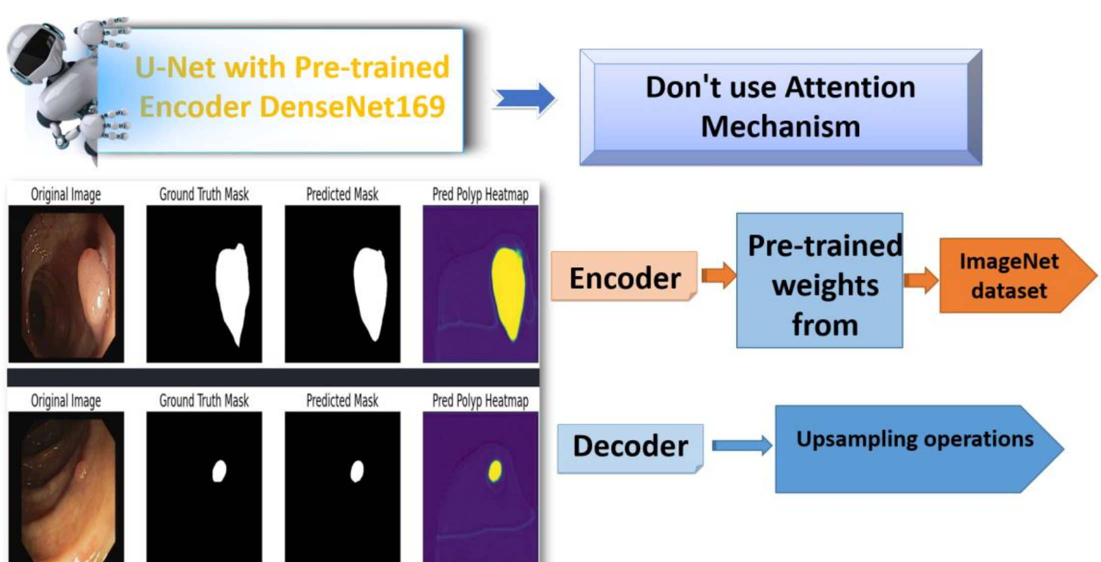


Fig.21. U-Net with Pre-trained Encoder DenseNet169

Machine and Deep learning Techniques for Histological and Endoscopic Images

Abdelouahab MOUSSAOUI, Computer Science Department
Faculty of Sciences, Ferhat Abbas University - Sétif 1, Algeria

Segmentation using SAM Model: With SAM, we can be using point or bounding boxes to pick the area of the image that is most interesting to us and extracting mask related to that area. it is can more interactive by using mouse to draw the bounding box around the area of interest.
After that we pass our bounding box trough the mask predictor to get our mask. the model gives us 3 predicted masks, but from those three the first two is the most important for us. Finally, the model generates the segmented image and display it.

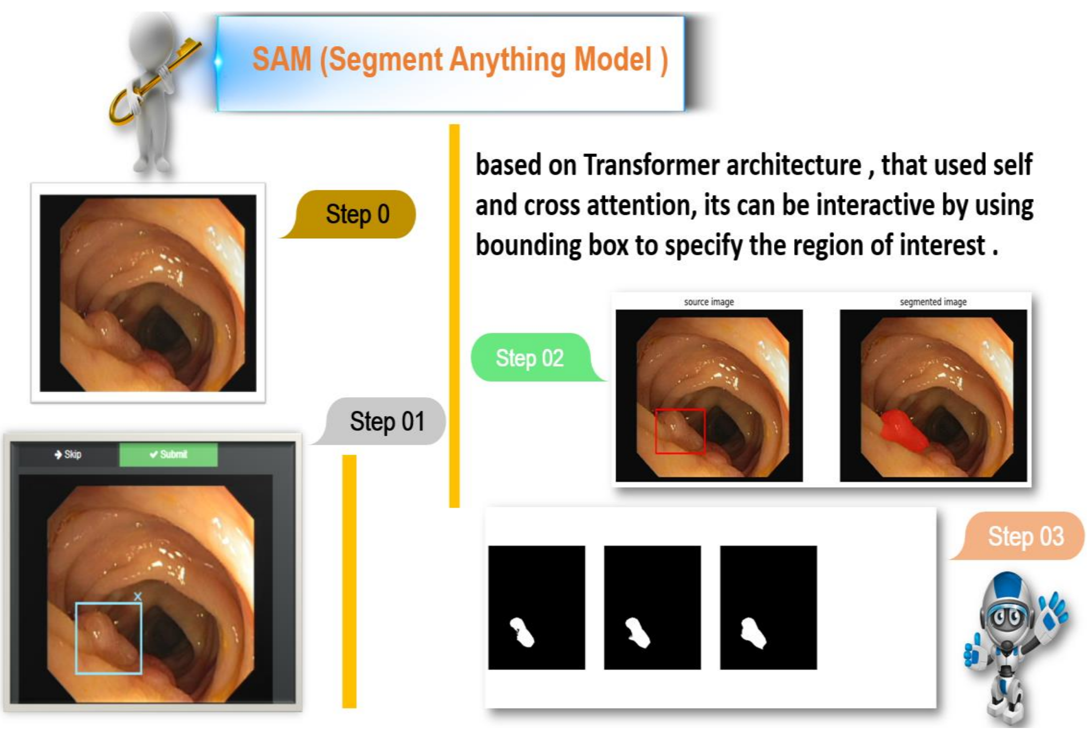


Fig.22. Case one of using SAM Model

Machine and Deep learning Techniques for Histological and Endoscopic Images

Abdelouahab MOUSSAOUI, Computer Science Department
Faculty of Sciences, Ferhat Abbas University - Sétif 1, Algeria

Also SAM can be used to segment visible colonic polyps from a dataset annotated with the bounding boxes and we will try to convert them into perfect segmentation masks which is obtained from Roboflow. we convert the format of the dataset into COCO format before integrated in the code, after we load annotation into the memory and print the list of classes, in our case we have 3 classes: Adenocarci noma, Adenoma, and Hyperplasic. finally, the model convert Image Bounding Box to segmented image and display the predicted masks. the model gives us 3 predicted masks, but from those three the first two is the most important for us.

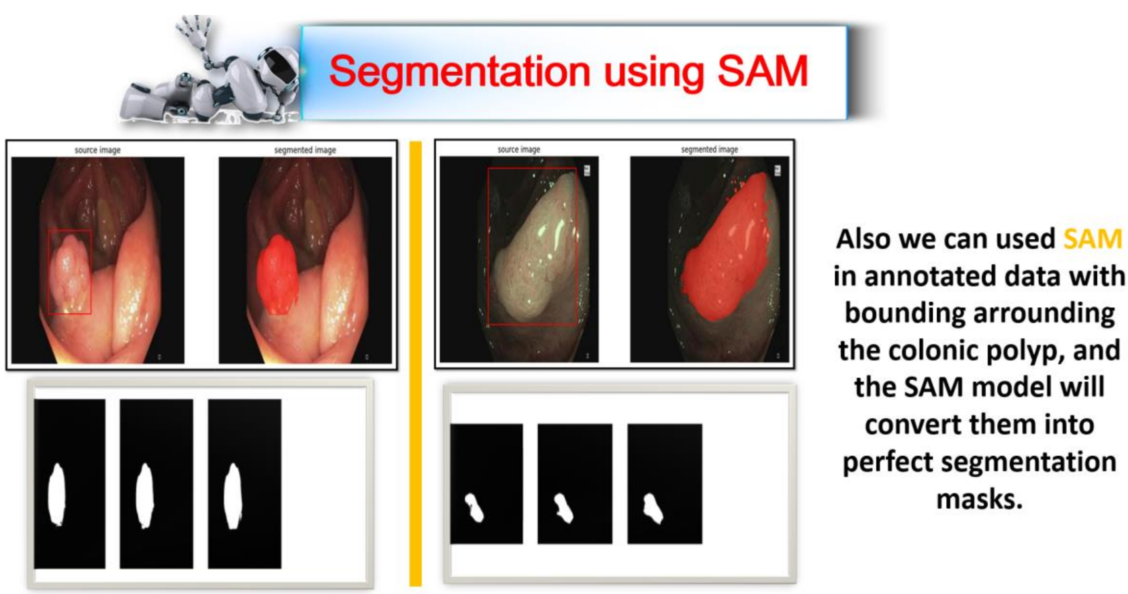


Fig.23. Case 2 of using SAM Model

Visual Computing Magazine

Machine and Deep learning Techniques for Histological and Endoscopic Images

Abdelouahab MOUSSAOUI, Computer Science Department
Faculty of Sciences, Ferhat Abbas University - Sétif 1, Algeria

Discussion Of Results

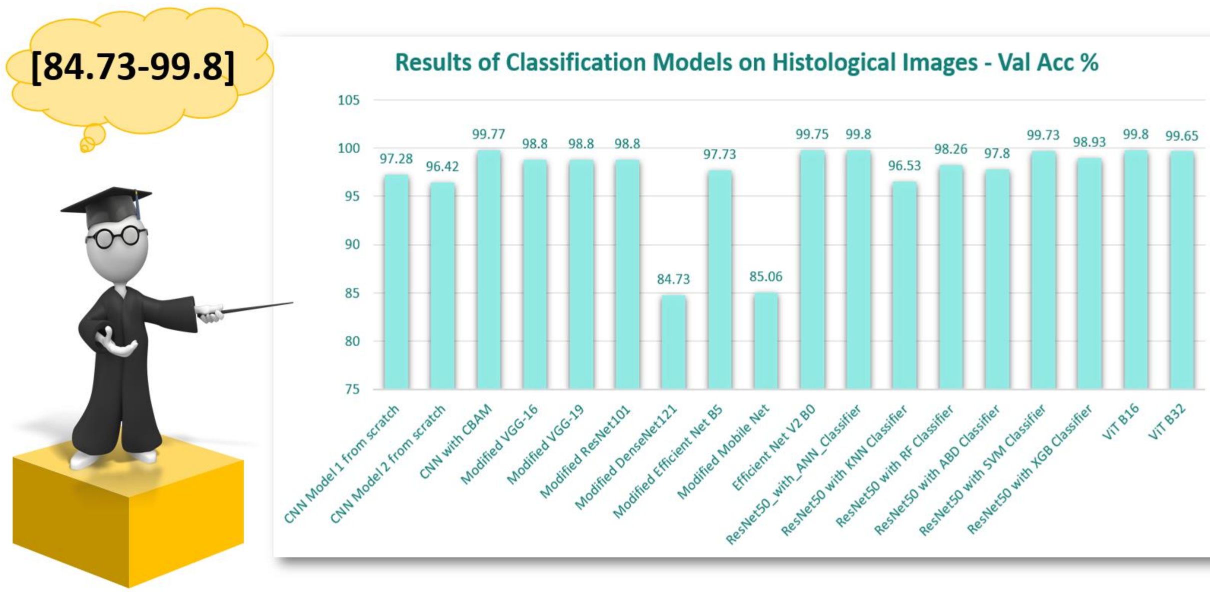


Fig.23. Case 2 of using SAM Model

The Main Model of Binary Classification of Colonic Tissues, is VIT B16 which have the biggest value of accuracy on validation-set, and other models are baseline models. The Main Model of Semantic Segmentation of Colonic Polyps, is FPN with Pre-Trained Encoder SE-ResNeXt50, and other models are baseline models.

Conclusion And Perspectives

This work represents case study and aims to develop and propose a new novel semantic segmentation and classification models, based on machine learning and deep learning to automatically segment colonic polyps better than the diagnosis of experts in pathology and manual-techniques, and classify colonic tissues for segmentation.

Machine and Deep learning Techniques for Histological and Endoscopic Images

Abdelouahab MOUSSAOUI, Computer Science Department
Faculty of Sciences, Ferhat Abbas University - Sétif 1, Algeria

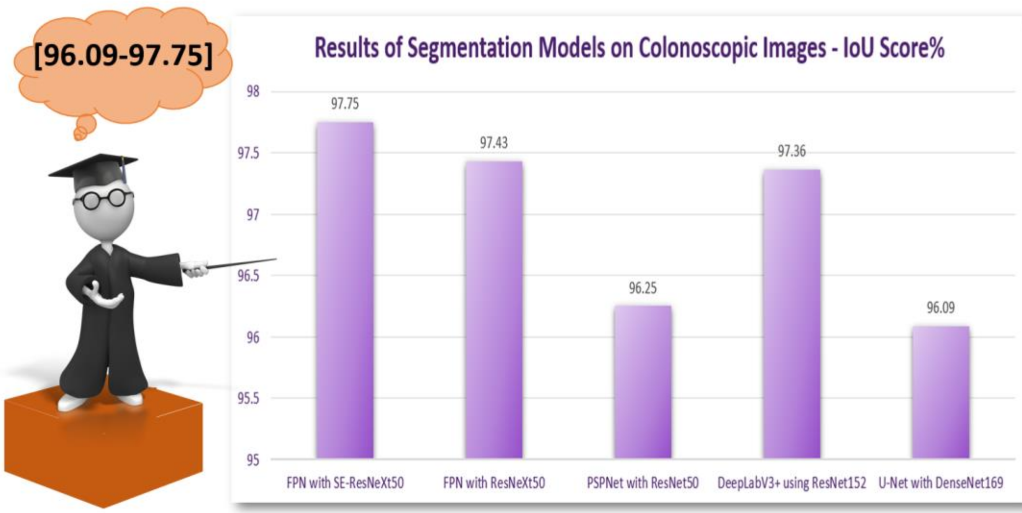


Fig.24. Results of Segmentation Models

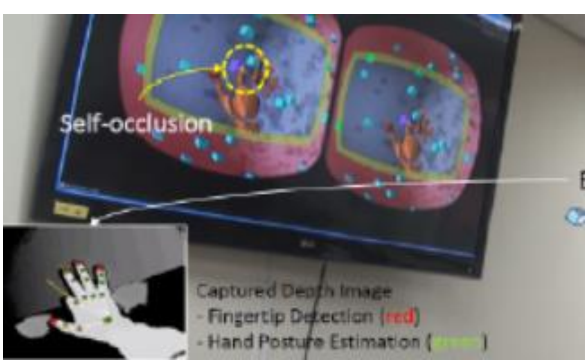
The purpose of our purposed models is to reduce the rate of false positive which occur when the model fails to detect or segment actual polyps present in the images, and for classification our proposed models provide more objective, standardized, and efficient methods for classification while reducing subjectivity and variability in the results.

For future, we would study further many related problems and test our models in different datasets to segment colonic polyps and classify colonic tissues, also we would combine a model for the segmentation of colonic polyps for endoscopic images with a model for the classification of colon tissue of histological images in one architecture, to do this is by using a multi-task learning approach, where both tasks (segmentation and classification) are learned jointly in a single model.

Visual Computing Magazine

Content

Vol 3, Issue 1, 2025



Visual Computing Magazine