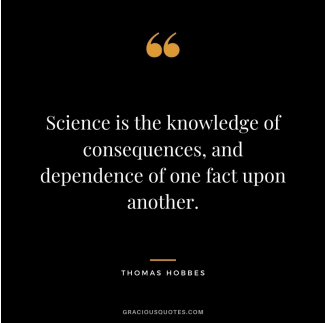


We introduced a new fully-neural approach for 3D human motion capture from monocular RGB videos with hard physics-based constraints which runs at interactive framerates and achieves state-of-the-art results on multiple metrics. Our neural physical model allows learning motion priors and the associated physical properties, as well as gain values of the neural PD controller from data. Thanks to the custom neural layer, which expresses hard physics-based constraints, our architecture is fully-differentiable. In addition, it can be trained jointly on several datasets thanks to the new form of input canonicalisation. Our experiments demonstrate that compared to PhysCap—a recent method with physics-based boundary conditions—our physical approach captures significantly faster motions, while being more accurate in terms of various 3D reconstruction metrics. Thanks to the full differentiability, the proposed method can be finetuned on datasets with 2D annotations only, which improves the reconstruction fidelity on in-the-wild footages. These properties make it well suitable for direct virtual character animation from monocular videos, without requiring any further post-processing of the estimated global 3D poses.

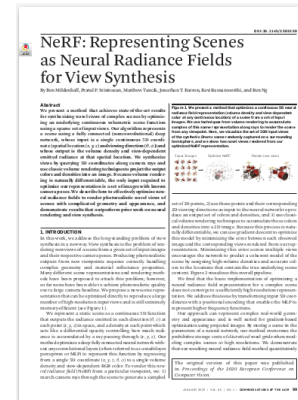


How to Read Academic Papers

Dr. Vladislav Golyanik, MPI for Informatics

Computer Vision and Machine Learning
for Computer Graphics
Seminar – Summer Term 2023

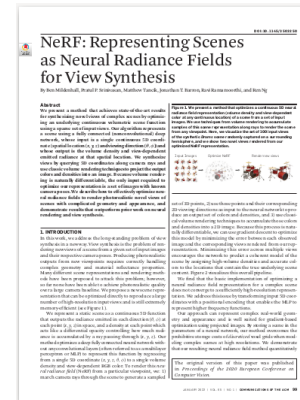
Scientific Papers vs Literary Fiction



Scientific Papers vs Literary Fiction



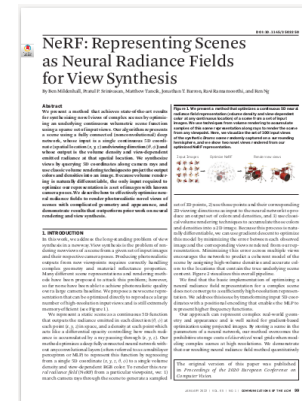
- 1) Tells a story
- 2) Covers wide range of topics
- 3) Written for general audience
- 4) Self-published (or by a publisher)
- 5) Less formal writing



Scientific Papers vs Literary Fiction



- 1) Tells a story
- 2) Covers wide range of topics
- 3) Written for general audience
- 4) Self-published (or by a publisher)
- 5) Less formal writing



- 1) Conveys scientific findings
- 2) Written to experts in the field
- 3) Uses technical language
- 4) Published in peer-reviewed venues
- 5) Certain structure is expected

The Primary Questions

Q: **Why** are papers published?

Q: **What** is the structure of papers?

Q: **How** to read academic papers?

The Primary Questions

Q: **Why** are papers published?

Q: **What** is the structure of papers?

Q: **How** to read academic papers?

A: ***It depends!***

The Primary Questions

Q: **Why** are papers published?

Q: **What** is the structure of papers?

Q: **How** to read academic papers?

A: ***It depends!***

The reasons why a paper is published influence its structure.
The structure influences how the paper is read and perceived.

Reasons to Publish

- Communication [of X] in a well structured form
- Documentation of work (math is the most precise language)
- Unpublished = Does not Exist
- Poor research should not be published

Reasons to Publish

- Communication [of X] in a well structured form
- Documentation of work (math is the most precise language)
- Unpublished = Does not Exist
- Poor research should not be published

X = {
new ideas, theories, algorithms, neural architectures
solutions to existing (e.g., long-standing) and new problems
combinations of components (existing and new)
current state of the art
opinion on a certain topic

Paper and Publication Types

Full paper

- Journal Article
- Conference Paper (Proceedings)
- Workshop Paper

Paper and Publication Types

Full paper

Journal Article

Conference Paper (Proceedings)

Workshop Paper



quality
predicted impact

Paper and Publication Types

Full paper

Journal Article
Conference Paper (Proceedings)
Workshop Paper

quality
predicted impact

nature



ICCV23

More types?

Paper and Publication Types

Full paper

Journal Article
Conference Paper (Proceedings)
Workshop Paper

quality
predicted impact



nature



ICCV23

More types?

Short paper

Survey/STAR

Opinion

Corrigendum

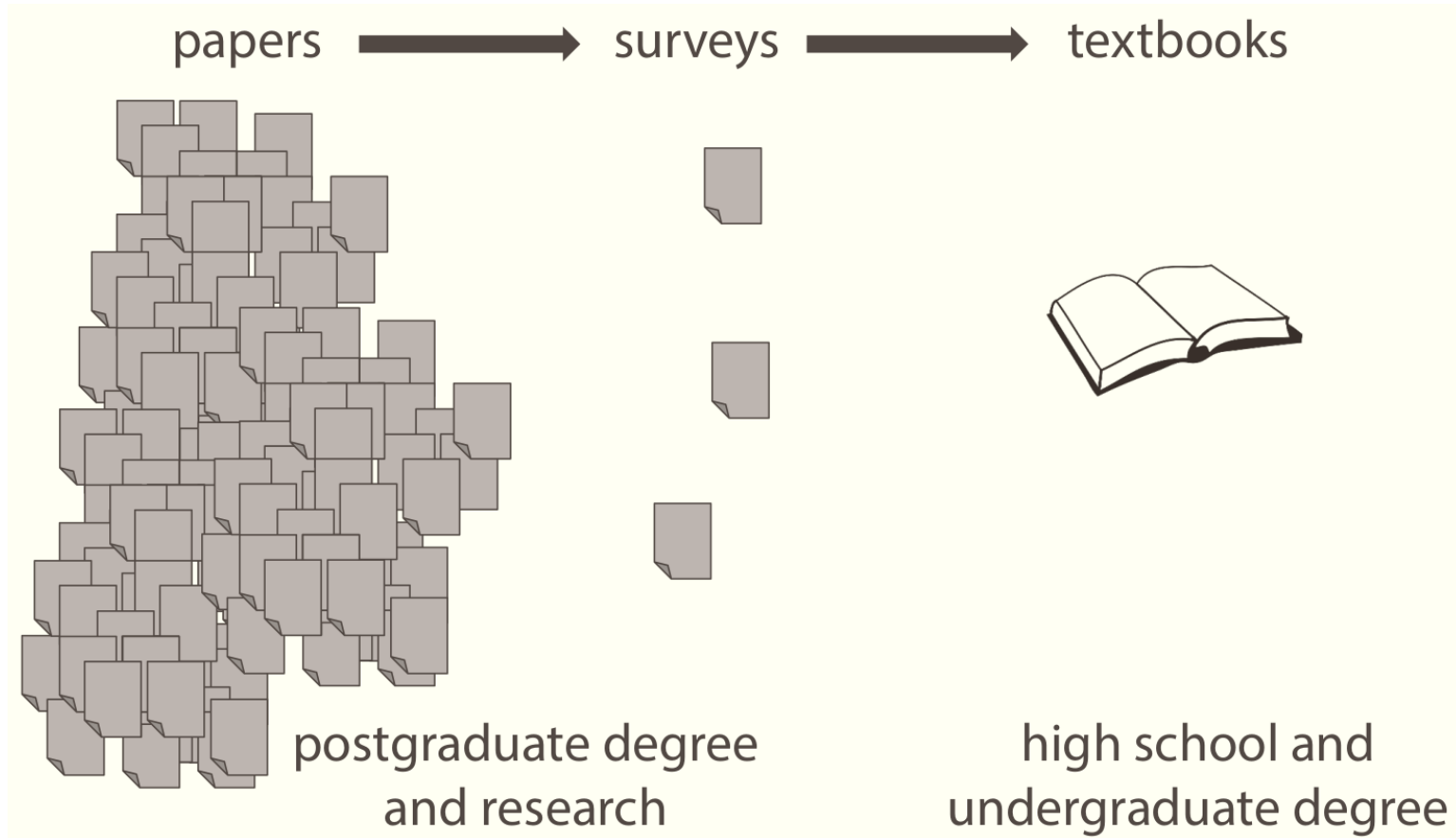
Technical Report (e.g., on arXiv)

Dissertation

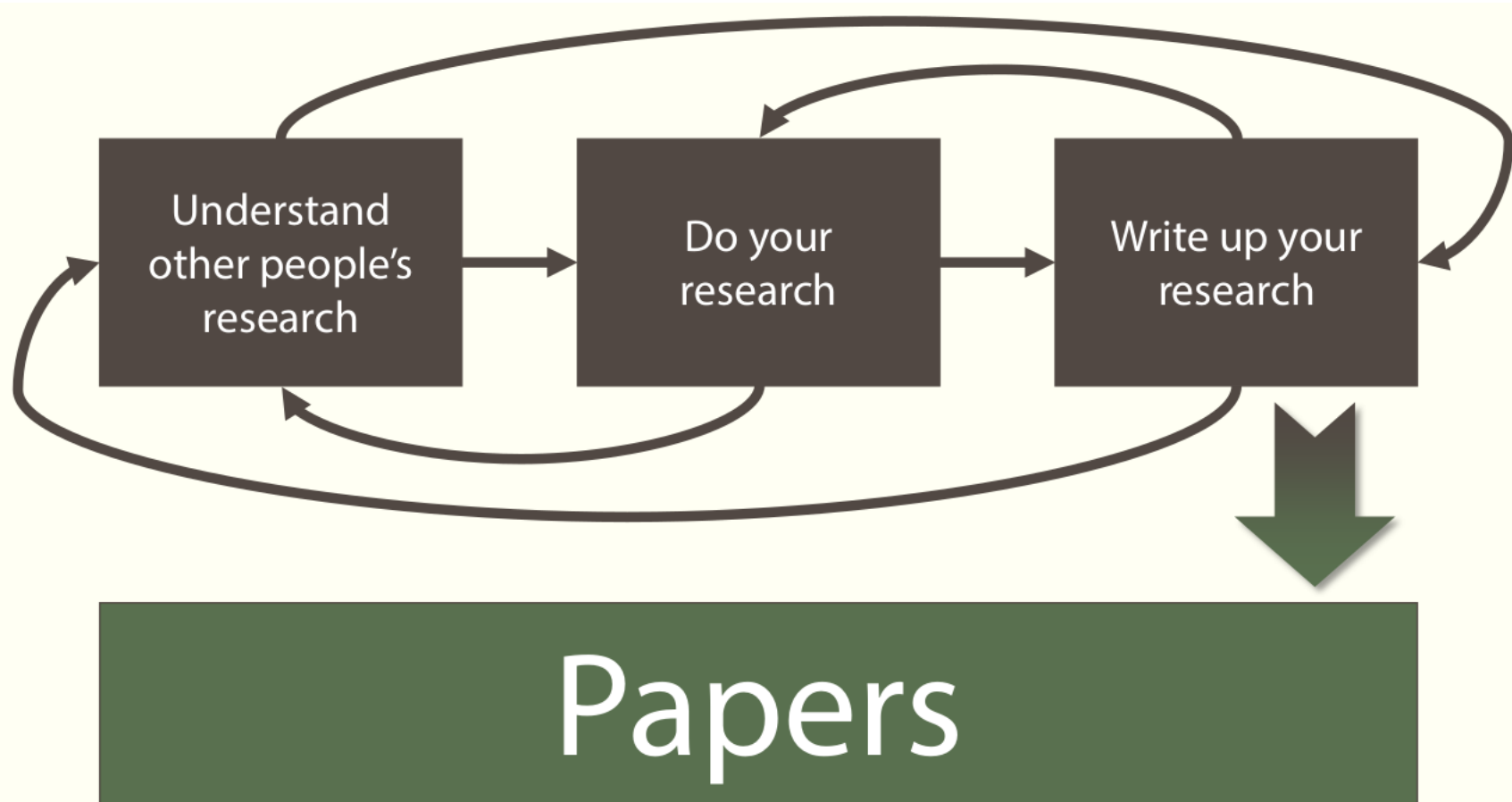
Book

Textbook

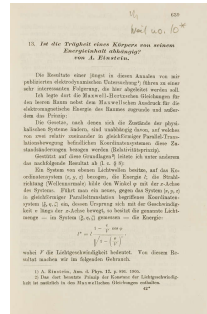
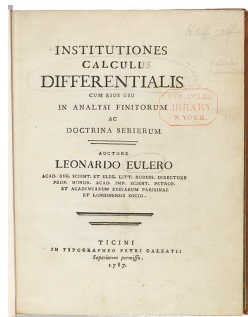
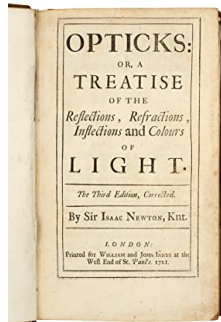
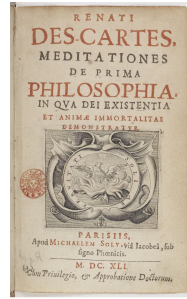
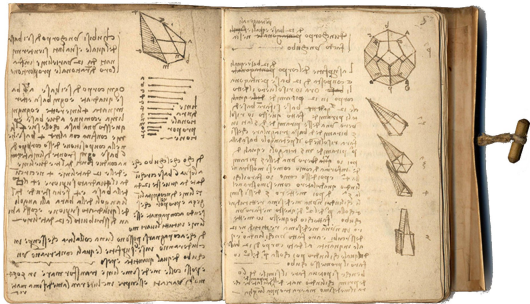
Academic Writing



Research and Papers



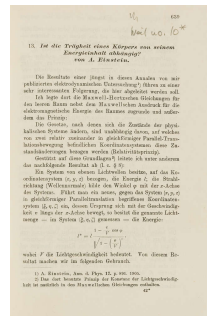
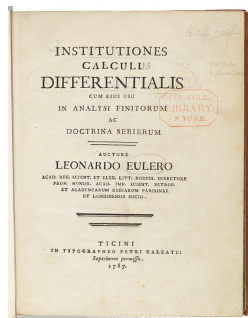
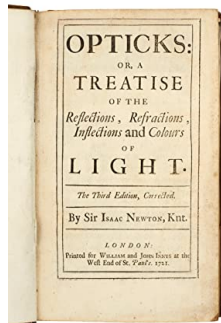
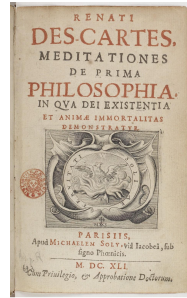
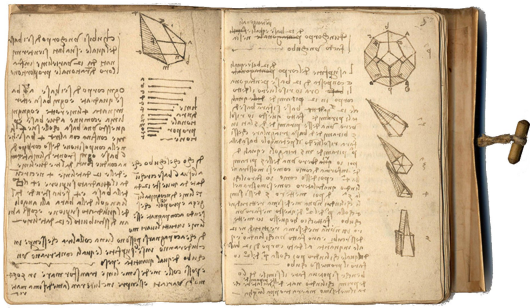
The Reviewing Process



back then

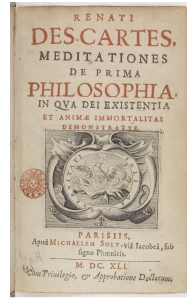
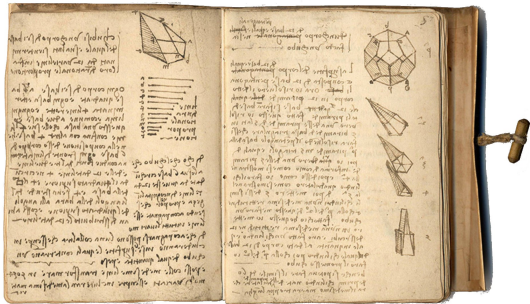
The Reviewing Process

2000+ papers at CVPR each year
~20k pages

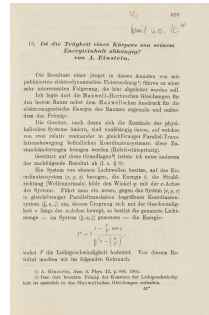
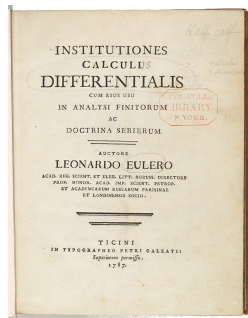
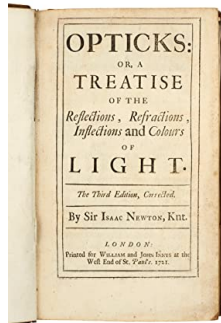


back then

The Reviewing Process



2000+ papers at CVPR each year
~20k pages



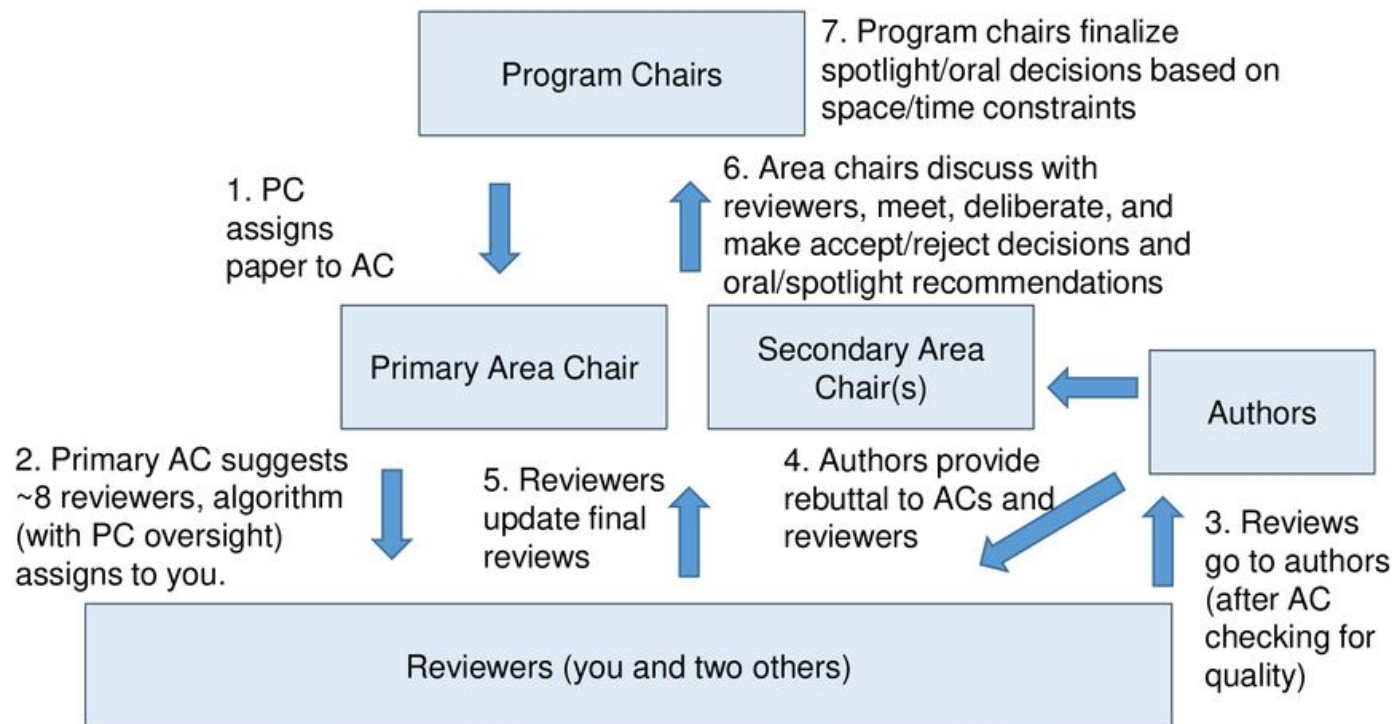
+

ACM SIGGRAPH, Eurographics, ICCV, ECCV, BMVC, GCPR, NeurIPS, ICLR, TPAMI, IJCV...

back then

The Reviewing Process

The decision process (overview)



Paper Structure

Neural Molecular ID Human Motion Capture with Physical Awareness

Abstract: This paper introduces a novel framework for capturing human motion with physical awareness. The proposed method, called Neural Molecular ID (NM-ID), leverages a deep learning architecture to learn a latent space representation of human motion. This latent space is used to generate high-quality motion capture data that is physically plausible and visually realistic. The framework is trained on a large dataset of human motion capture data and is able to generalize to new motion capture scenarios. The results show that NM-ID outperforms existing methods in terms of motion quality and physical awareness.

Figure 1: Visualizations of motion capture data. The top row shows raw motion capture data for a person walking. The bottom row shows the corresponding motion capture data generated by the proposed method, which appears smoother and more physically realistic.

Methodology and Results

The proposed framework consists of a neural network that takes input motion capture data and outputs a latent space representation. This latent space is used to generate high-quality motion capture data that is physically plausible and visually realistic. The framework is trained on a large dataset of human motion capture data and is able to generalize to new motion capture scenarios. The results show that the proposed method outperforms existing methods in terms of motion quality and physical awareness.

Figure 2: Comparison of motion capture data. The top row shows raw motion capture data. The bottom row shows the corresponding motion capture data generated by the proposed method. The proposed method's results are smoother and more physically realistic.

Conclusion and Future Work

The proposed framework provides a novel approach to capturing human motion with physical awareness. The results show that the proposed method outperforms existing methods in terms of motion quality and physical awareness. Future work includes extending the framework to capture more complex motions and improving the physical awareness of the generated motion capture data.

Appendix

Figure 3: Comparison of motion capture data. The top row shows raw motion capture data. The bottom row shows the corresponding motion capture data generated by the proposed method. The proposed method's results are smoother and more physically realistic.

References

1. [1] Smith, J., and Jones, K. "A study of human motion capture." *Journal of Computer Graphics*, 2018.

2. [2] Brown, L., and Green, M. "A study of human motion capture." *Journal of Computer Graphics*, 2019.

3. [3] White, R., and Black, N. "A study of human motion capture." *Journal of Computer Graphics*, 2020.

Methodology and Results

The proposed framework consists of a neural network that takes input motion capture data and outputs a latent space representation. This latent space is used to generate high-quality motion capture data that is physically plausible and visually realistic. The framework is trained on a large dataset of human motion capture data and is able to generalize to new motion capture scenarios. The results show that the proposed method outperforms existing methods in terms of motion quality and physical awareness.

Methodology and Results

The proposed framework consists of a neural network that takes input motion capture data and outputs a latent space representation. This latent space is used to generate high-quality motion capture data that is physically plausible and visually realistic. The framework is trained on a large dataset of human motion capture data and is able to generalize to new motion capture scenarios. The results show that the proposed method outperforms existing methods in terms of motion quality and physical awareness.

Methodology and Results

The proposed framework consists of a neural network that takes input motion capture data and outputs a latent space representation. This latent space is used to generate high-quality motion capture data that is physically plausible and visually realistic. The framework is trained on a large dataset of human motion capture data and is able to generalize to new motion capture scenarios. The results show that the proposed method outperforms existing methods in terms of motion quality and physical awareness.

Methodology and Results

The proposed framework consists of a neural network that takes input motion capture data and outputs a latent space representation. This latent space is used to generate high-quality motion capture data that is physically plausible and visually realistic. The framework is trained on a large dataset of human motion capture data and is able to generalize to new motion capture scenarios. The results show that the proposed method outperforms existing methods in terms of motion quality and physical awareness.

Methodology and Results

The proposed framework consists of a neural network that takes input motion capture data and outputs a latent space representation. This latent space is used to generate high-quality motion capture data that is physically plausible and visually realistic. The framework is trained on a large dataset of human motion capture data and is able to generalize to new motion capture scenarios. The results show that the proposed method outperforms existing methods in terms of motion quality and physical awareness.

Methodology and Results

The proposed framework consists of a neural network that takes input motion capture data and outputs a latent space representation. This latent space is used to generate high-quality motion capture data that is physically plausible and visually realistic. The framework is trained on a large dataset of human motion capture data and is able to generalize to new motion capture scenarios. The results show that the proposed method outperforms existing methods in terms of motion quality and physical awareness.

Methodology and Results

The proposed framework consists of a neural network that takes input motion capture data and outputs a latent space representation. This latent space is used to generate high-quality motion capture data that is physically plausible and visually realistic. The framework is trained on a large dataset of human motion capture data and is able to generalize to new motion capture scenarios. The results show that the proposed method outperforms existing methods in terms of motion quality and physical awareness.

Methodology and Results

The proposed framework consists of a neural network that takes input motion capture data and outputs a latent space representation. This latent space is used to generate high-quality motion capture data that is physically plausible and visually realistic. The framework is trained on a large dataset of human motion capture data and is able to generalize to new motion capture scenarios. The results show that the proposed method outperforms existing methods in terms of motion quality and physical awareness.

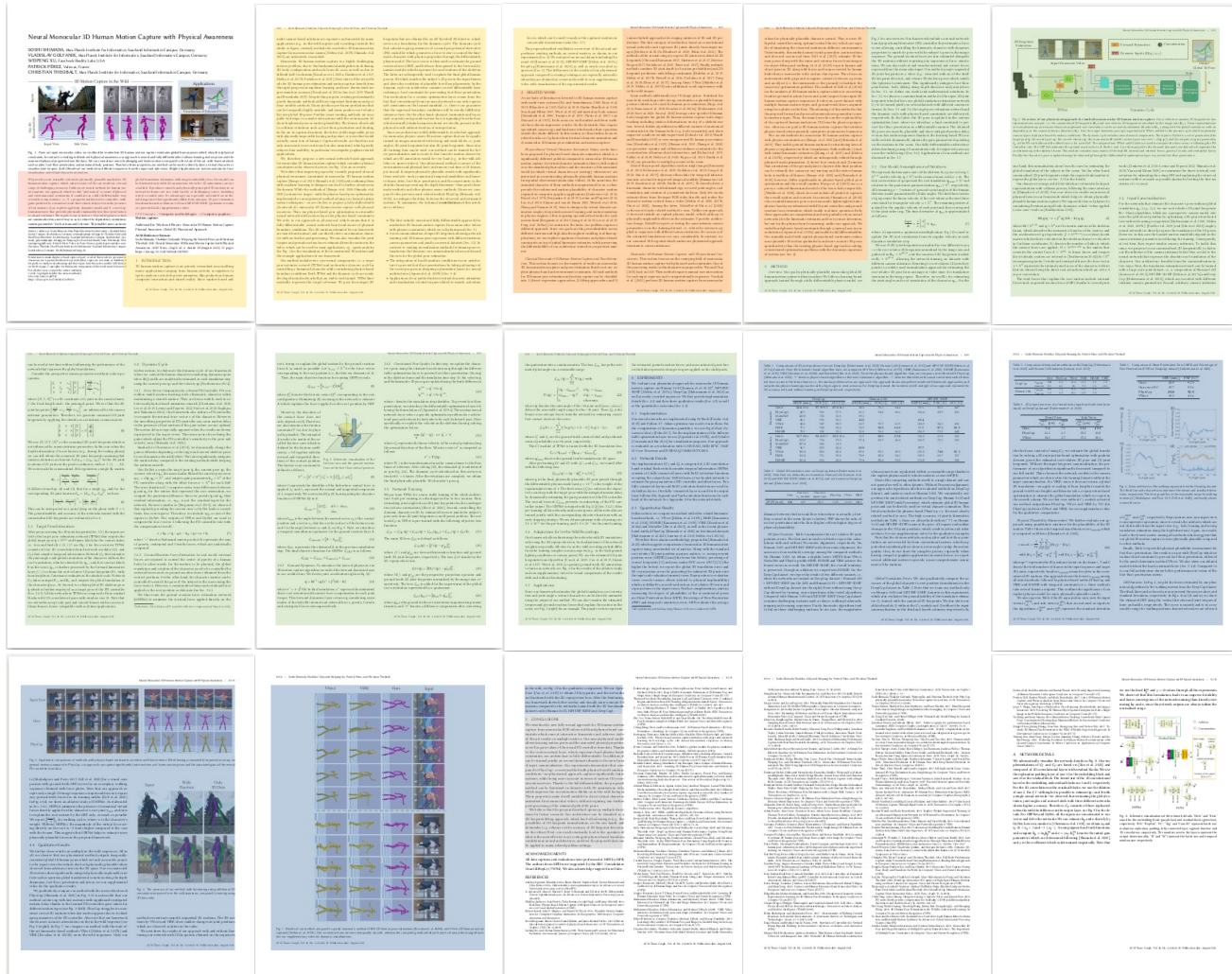
Methodology and Results

The proposed framework consists of a neural network that takes input motion capture data and outputs a latent space representation. This latent space is used to generate high-quality motion capture data that is physically plausible and visually realistic. The framework is trained on a large dataset of human motion capture data and is able to generalize to new motion capture scenarios. The results show that the proposed method outperforms existing methods in terms of motion quality and physical awareness.

Methodology and Results

The proposed framework consists of a neural network that takes input motion capture data and outputs a latent space representation. This latent space is used to generate high-quality motion capture data that is physically plausible and visually realistic. The framework is trained on a large dataset of human motion capture data and is able to generalize to new motion capture scenarios. The results show that the proposed method outperforms existing methods in terms of motion quality and physical awareness.

Paper Structure



Paper Structure

Title / Header

Abstract

1. Introduction

2. Related Work

3. Method

4. Experiments

5. Conclusions

Acknowledgements

References

Appendix



Paper Structure

Title / Header

Abstract

1. Introduction

2. Related Work

3. Method

4. Experiments

5. Conclusions

Acknowledgements

References

Appendix

Video Poster

Webpage

Source Code

The image displays a grid of 16 panels representing the structure of a research paper. The central panel is a video poster for the paper "Neural Monocular 3D Human Motion Capture with Physical Awareness" by Soshi Shimada, Vladislav Golyanik, Weipeng Xu, Patrick Pérez, and Christian Theobalt. The poster features a SIGGRAPH 2021 logo, a 3D visualization of a person in a virtual environment, and logos for Max Planck Institute for Informatics, Saarland Informatics Campus, facebook Reality Labs, and Valco.ai. The surrounding panels show various parts of the paper, including abstracts, introductions, methods, experiments, and conclusions.

Abstract

Marker-less 3D human motion capture from a single colour camera has seen significant progress. However, it is a very challenging and severely ill-posed problem. In consequence, even the most accurate state-of-the-art approaches have significant limitations. Purely kinematic formulations on the basis of individual joints or skeletons, and the frequent frame-wise reconstruction in state-of-the-art methods greatly limit 3D accuracy and temporal stability compared to multi-view or marker-based motion capture. Further, captured 3D poses are often physically incorrect and biomechanically implausible, or exhibit implausible environment interactions (floor penetration, foot skating, unnatural body leaning and strong shifting in depth), which is problematic for any use case in computer graphics.

We, therefore, present *PhysCap*, the first algorithm for physically plausible, real-time and marker-less human 3D motion capture with a single colour camera at 25 fps. Our algorithm first captures 3D human poses purely kinematically. To this end, a CNN infers 2D and 3D joint positions, and subsequently, an inverse kinematics step finds space-time coherent joint angles and global 3D pose. Next, these kinematic reconstructions are used as constraints in a real-time physics-based pose optimiser that accounts for environment constraints (e.g., collision handling and floor placement), gravity, and biophysical plausibility of human postures. Our approach employs a combination of ground reaction force and residual force for plausible root control, and uses a trained neural network to detect foot contact events in images. Our method captures physically plausible and temporally stable global 3D human motion, without physically implausible postures, floor penetrations or foot skating, from video in real time and in general scenes. *PhysCap* achieves state-of-the-art accuracy on established pose benchmarks, and we propose new metrics to demonstrate the improved physical plausibility and temporal stability.

Abstract

Marker-less 3D human motion capture from a single colour camera has seen significant progress. However, it is a very challenging and severely ill-posed problem. In consequence, even the most accurate state-of-the-art approaches have significant limitations. Purely kinematic formulations on the basis of individual joints or skeletons, and the frequent frame-wise reconstruction in state-of-the-art methods greatly limit 3D accuracy and temporal stability compared to multi-view or marker-based motion capture. Further, captured 3D poses are often physically incorrect and biomechanically implausible, or exhibit implausible environment interactions (floor penetration, foot skating, unnatural body leaning and strong shifting in depth), which is problematic for any use case in computer graphics.

We, therefore, present *PhysCap*, the first algorithm for physically plausible, real-time and marker-less human 3D motion capture with a single colour camera at 25 fps. Our algorithm first captures 3D human poses purely kinematically. To this end, a CNN infers 2D and 3D joint positions, and subsequently, an inverse kinematics step finds space-time coherent joint angles and global 3D pose. Next, these kinematic reconstructions are used as constraints in a real-time physics-based pose optimiser that accounts for environment constraints (e.g., collision handling and floor placement), gravity, and biophysical plausibility of human postures. Our approach employs a combination of ground reaction force and residual force for plausible root control, and uses a trained neural network to detect foot contact events in images. Our method captures physically plausible and temporally stable global 3D human motion, without physically implausible postures, floor penetrations or foot skating, from video in real time and in general scenes. *PhysCap* achieves state-of-the-art accuracy on established pose benchmarks, and we propose new metrics to demonstrate the improved physical plausibility and temporal stability.



problem



method/contributions



challenges



experimental set-up and results

Abstract

Motion segmentation is a challenging problem that seeks to identify independent motions in two or several input images. This paper introduces the first algorithm for motion segmentation that relies on adiabatic quantum optimization of the objective function. The proposed method achieves on-par performance with the state of the art on problem instances which can be mapped to modern quantum annealers.



problem



method/contributions



challenges

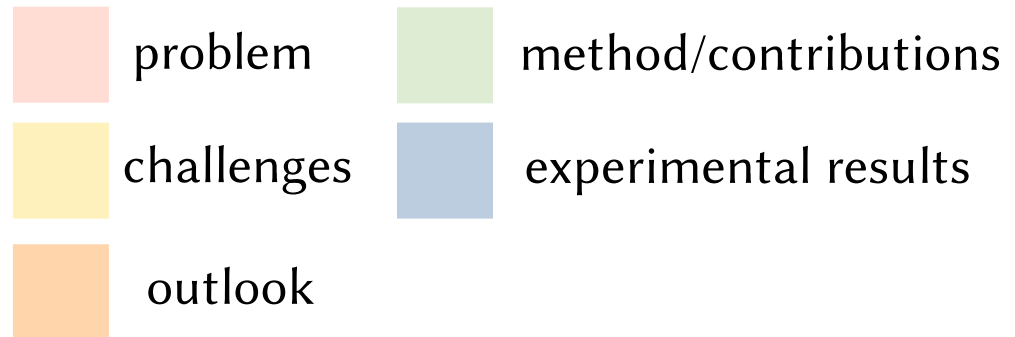


experimental set-up and results

Conclusions

We introduced a new fully-neural approach for 3D human motion capture from monocular RGB videos with hard physics-based constraints which runs at interactive framerates and achieves state-of-the-art results on multiple metrics. Our neural physical model allows learning motion priors and the associated physical properties, as well as gain values of the neural PD controller from data. Thanks to the custom neural layer, which expresses hard physics-based constraints, our architecture is fully-differentiable. In addition, it can be trained jointly on several datasets thanks to the new form of input canonicalisation. Our experiments demonstrate that compared to PhysCap—a recent method with physics-based boundary conditions—our physionical approach captures significantly faster motions, while being more accurate in terms of various 3D reconstruction metrics. Thanks to the full differentiability, the proposed method can be finetuned on datasets with 2D annotations only, which improves the reconstruction fidelity on in-the-wild footages. These properties make it well suitable for direct virtual character animation from monocular videos, without requiring any further post-processing of the estimated global 3D poses.

We believe that the proposed method opens up multiple directions for future research. Our architecture can be classified as a 2D keypoint lifting approach, which has both advantages (*e.g.*, the possibility of 2D keypoint normalisation, on the one hand) and downsides (*e.g.*, reliance on the accuracy of 2D keypoint detectors, on the other). Next, our results naturally lead to the question of what is the most effective way to integrate physics-based boundary conditions in neural architectures, and how the proposed ideas can be applied to many related problem settings.



Scientific Writing

Main principles:

Objectivity

Precision

Clarity

Efficiency

Scientific Writing

Main principles:

Objectivity
Precision
Clarity
Efficiency

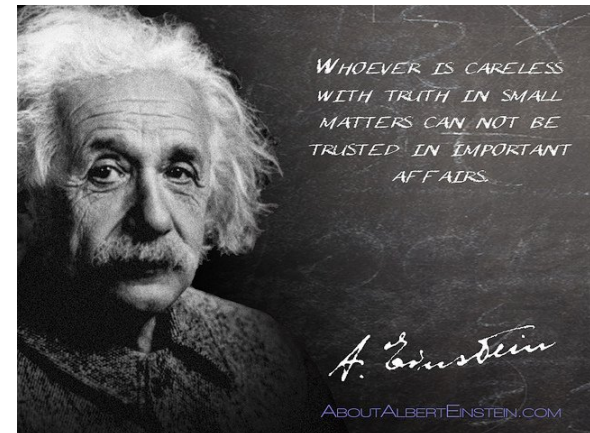
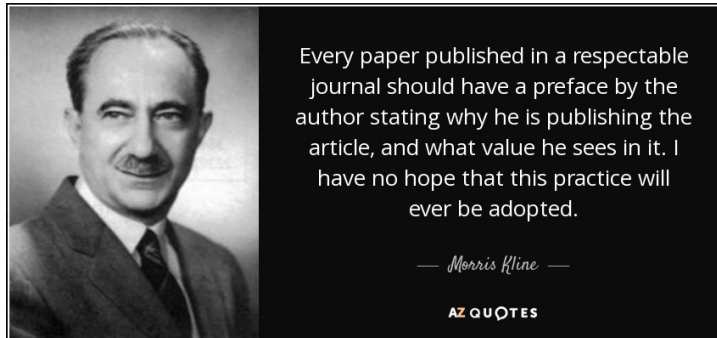
- Each discipline follows its **set of rules, conventions and best practices**
- Focus is on **information**
- Scientific arguments are built **solely on evidence and logic** and do not include emotions or opinions
- Scientists want their readers **to draw the same conclusions** from the evidence that they did; they, therefore, must present their chain of logic as clearly as possible
- Readers want to be able **to easily evaluate the validity of results** and conclusions, using the evidence they have before them
- All sources must be **cited**

Questions to Ask While Reading

- What is the paper trying to convey?
- Why are the research and the obtained results significant?
- How were the results evaluated/measured?
- What were the results?
- What is the conclusion, and why?
- Do I trust the findings?

Questions to Ask While Reading

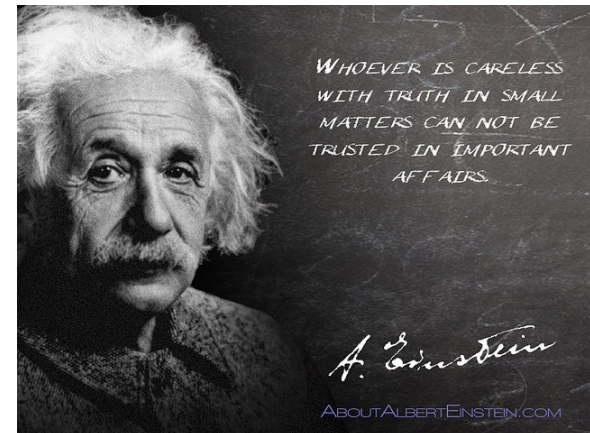
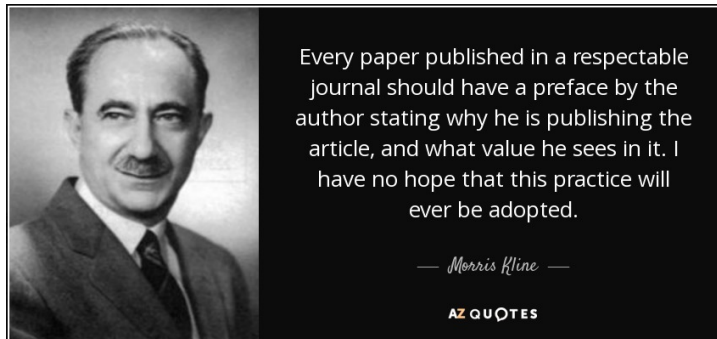
- What is the paper trying to convey?
- Why are the research and the obtained results significant?
- How were the results evaluated/measured?
- What were the results?
- What is the conclusion, and why?
- Do I trust the findings?



Questions to Ask While Reading

- What is the paper trying to convey?
- Why are the research and the obtained results significant?
- How were the results evaluated/measured?
- What were the results?
- What is the conclusion, and why?
- Do I trust the findings?

Be Critical!
Ask questions!



How to Read a Technical Paper

Q: What is your goal (when to stop)?

How to Read a Technical Paper

Q: What is your goal (when to stop)?

→ To see qualitative results

skim through the paper

→ To learn what it is about

+ read Abstract, Conclusions and Discussion

→ To understand the main idea

+ read Introduction

→ To understand most details

+ read Method and Experimental sections

* To understand in detail how it relates to previous methods

+ read Related Work

The Three-Pass Approach

How to Read a Paper

S. Keshav

David R. Cheriton School of Computer Science, University of Waterloo
Waterloo, ON, Canada
keshav@uwaterloo.ca

ABSTRACT

Researchers spend a great deal of time reading research papers. However, this skill is rarely taught, leading to much wasted effort. This article outlines a practical and efficient *three-pass method* for reading research papers. I also describe how to use this method to do a literature survey.

Categories and Subject Descriptors: A.1 [Introductory and Survey]

General Terms: Documentation.

Keywords: Paper, Reading, Hints.

1. INTRODUCTION

Researchers must read papers for several reasons: to review them for a conference or a class, to keep current in their field, or for a literature survey of a new field. A typical researcher will likely spend hundreds of hours every year

4. Glance over the references, mentally ticking off the ones you've already read

At the end of the first pass, you should be able to answer the *five Cs*:

1. *Category*: What type of paper is this? A measurement paper? An analysis of an existing system? A description of a research prototype?
2. *Context*: Which other papers is it related to? Which theoretical bases were used to analyze the problem?
3. *Correctness*: Do the assumptions appear to be valid?
4. *Contributions*: What are the paper's main contributions?
5. *Clarity*: Is the paper well written?

The Three-Pass Approach

2.1 The first pass

The first pass is a quick scan to get a bird's-eye view of the paper. You can also decide whether you need to do any more passes. This pass should take about five to ten minutes and consists of the following steps:

1. Carefully read the title, abstract, and introduction
2. Read the section and sub-section headings, but ignore everything else
3. Read the conclusions
4. Glance over the references, mentally ticking off the ones you've already read

2.2 The second pass

In the second pass, read the paper with greater care, but ignore details such as proofs. It helps to jot down the key points, or to make comments in the margins, as you read.

1. Look carefully at the figures, diagrams and other illustrations in the paper. Pay special attention to graphs. Are the axes properly labeled? Are results shown with error bars, so that conclusions are statistically significant? Common mistakes like these will separate rushed, shoddy work from the truly excellent.
2. Remember to mark relevant unread references for further reading (this is a good way to learn more about the background of the paper).

The Three-Pass Approach

2.3 The third pass

To fully understand a paper, particularly if you are reviewer, requires a third pass. The key to the third pass is to attempt to *virtually re-implement* the paper: that is, making the same assumptions as the authors, re-create the work. By comparing this re-creation with the actual paper, you can easily identify not only a paper's innovations, but also its hidden failings and assumptions.

This pass requires great attention to detail. You should identify and challenge every assumption in every statement. Moreover, you should think about how you yourself would present a particular idea. This comparison of the actual with the virtual lends a sharp insight into the proof and presentation techniques in the paper and you can very likely add this to your repertoire of tools. During this pass, you should also jot down ideas for future work.

This pass can take about four or five hours for beginners, and about an hour for an experienced reader. At the end of this pass, you should be able to reconstruct the entire structure of the paper from memory, as well as be able to identify its strong and weak points. In particular, you should be able to pinpoint implicit assumptions, missing citations to relevant work, and potential issues with experimental or analytical techniques.

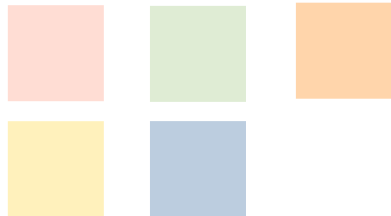
- Attempt to virtually re-implement the paper
- Requires high attention to detail
- Enables identifying strong and weak points
- Takes up to multiple hours

Remember What You Read

- Make notes while reading papers
- Keep track of papers in a written form (title, authors, venue, link, the main idea)
- Write a summary of the most relevant papers
- Use reference managers

Conclusion

- Papers convey scientific findings and are written for experts
- Papers differ in their type and quality
- Published papers are peer-reviewed
- Papers have a predefined (conventional) structure
- Principles of scientific writing: objectivity, precision, clarity, efficiency
- The three-pass approach
- How to read a paper depends on the goal



Be Critical!
Ask questions!

THE THREE-PASS APPROACH

The key idea is that you should read the paper in up to three passes, instead of starting at the beginning and plowing your way to the end. Each pass accomplishes specific goals and builds upon the previous pass: The *first* pass gives you a general idea about the paper. The *second* pass lets you grasp the paper's content, but not its details. The *third* pass helps you understand the paper in depth.

Questions?