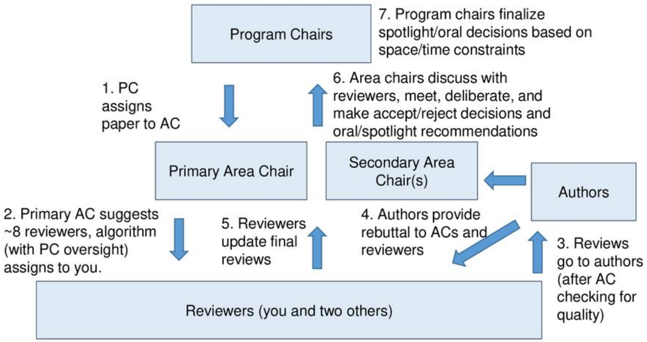


Program Chairs

7. Program chairs finalize spotlight/oral decisions based on space/time constraints

1. PC assigns paper to AC

6. Area chairs discuss with reviewers, meet, deliberate, and make accept/reject decisions and oral/spotlight recommendations

Primary Area Chair

Secondary Area Chair(s)

Authors

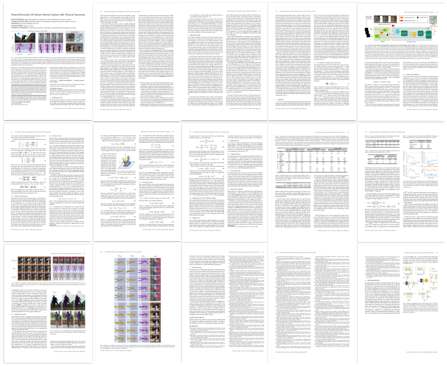
2. Primary AC suggests ~8 reviewers, algorithm (with PC oversight) assigns to you.

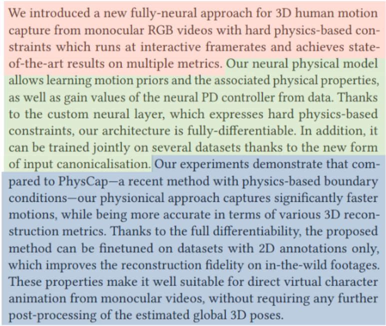
5. Reviewers update final reviews

4. Authors provide rebuttal to ACs and reviewers

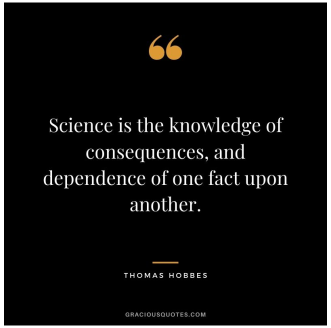
3. Reviews go to authors (after AC checking for quality)

Reviewers (you and two others)





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.



“

Science is the knowledge of consequences, and dependence of one fact upon another.

THOMAS HOBBES

GRACIOUSQUOTES.COM

How to Read Academic Papers

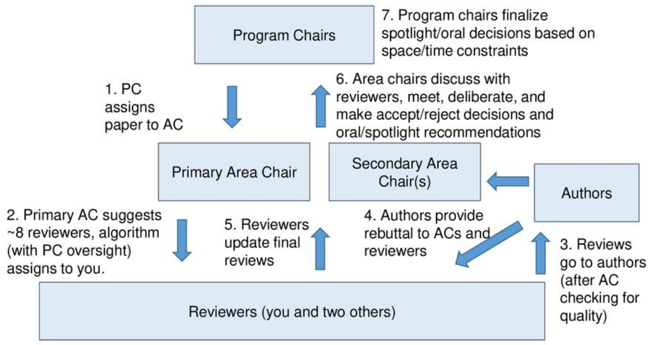
Dr. Vladislav Golyanik, MPI for Informatics

Classical Concepts of Computer Vision and Computer Graphics in the Neural Age

Seminar – Summer Term 2024

max planck institut informatik

Overview



Program Chairs

7. Program chairs finalize spotlight/oral decisions based on space/time constraints

1. PC assigns paper to AC

6. Area chairs discuss with reviewers, meet, deliberate, and make accept/reject decisions and oral/spotlight recommendations

Primary Area Chair

Secondary Area Chair(s)

Authors

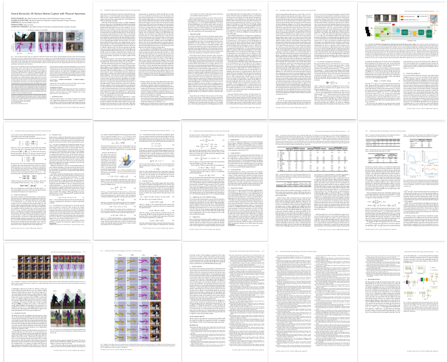
2. Primary AC suggests ~8 reviewers, algorithm (with PC oversight) assigns to you.

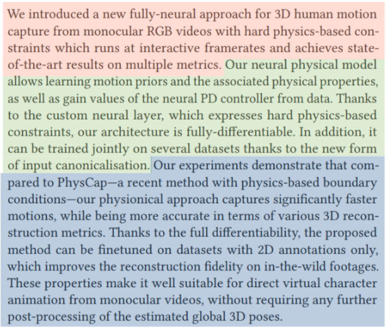
5. Reviewers update final reviews

4. Authors provide rebuttal to ACs and reviewers

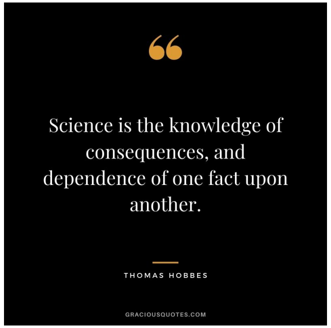
3. Reviews go to authors (after AC checking for quality)

Reviewers (you and two others)





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.

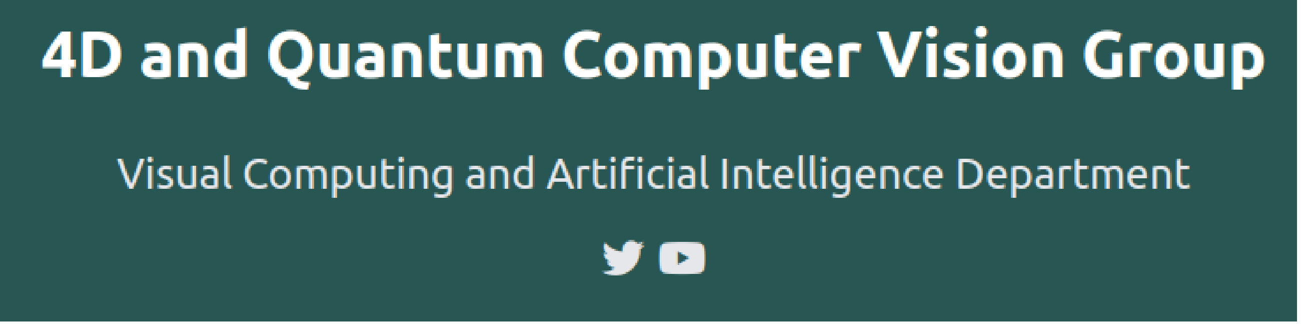


Science is the knowledge of consequences, and dependence of one fact upon another.

THOMAS HOBBES

GRACIOUSQUOTES.COM

- Reasons to Publish
- Scientific Papers and Their Types
- The Reviewing Process
- Paper Structure
- How to Read a Technical Paper
- The Three-Pass Approach



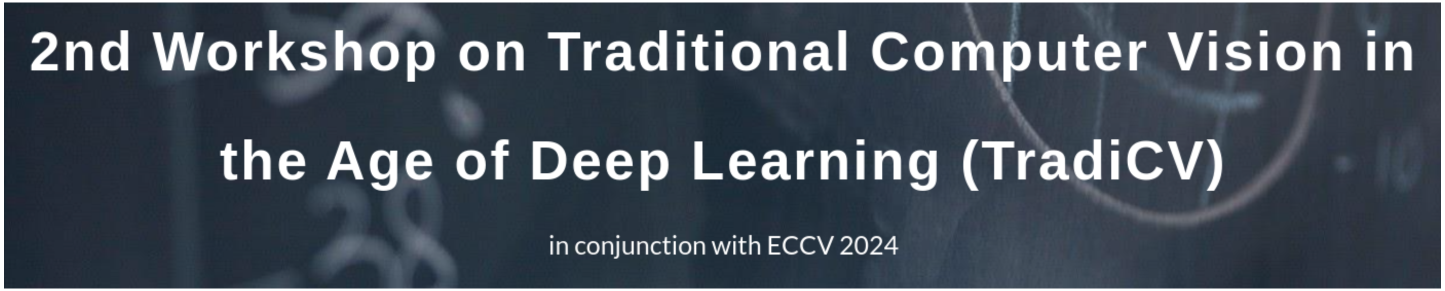
4D and Quantum Computer Vision Group

Visual Computing and Artificial Intelligence Department

4D and Quantum Vision Group

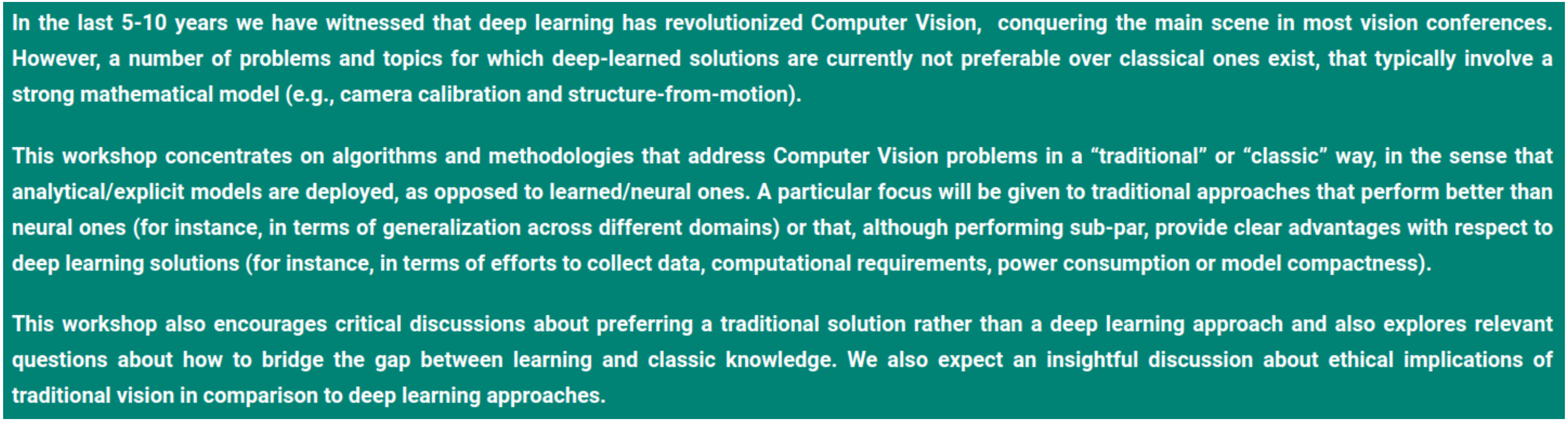
- Today’s Instructor: Dr. Vladislav Golyanik
- Affiliation: 4DQV at VCAI Dep./ MPI for Informatics
- Contact: golyanik@mpi-inf.mpg.de
- Research Interests:
  - 3D/4D Reconstruction and Neural Rendering
  - 3D Generative Models
  - Quantum Computer Vision

Photo taken by Marion Fregeac



2nd Workshop on Traditional Computer Vision in the Age of Deep Learning (TradiCV)

in conjunction with ECCV 2024



In the last 5-10 years we have witnessed that deep learning has revolutionized Computer Vision, conquering the main scene in most vision conferences. However, a number of problems and topics for which deep-learned solutions are currently not preferable over classical ones exist, that typically involve a strong mathematical model (e.g., camera calibration and structure-from-motion).

This workshop concentrates on algorithms and methodologies that address Computer Vision problems in a “traditional” or “classic” way, in the sense that analytical/explicit models are deployed, as opposed to learned/neural ones. A particular focus will be given to traditional approaches that perform better than neural ones (for instance, in terms of generalization across different domains) or that, although performing sub-par, provide clear advantages with respect to deep learning solutions (for instance, in terms of efforts to collect data, computational requirements, power consumption or model compactness).

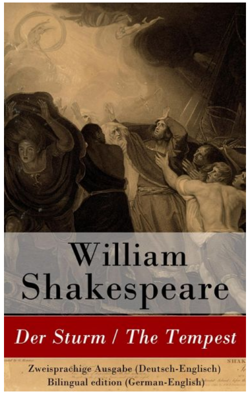
This workshop also encourages critical discussions about preferring a traditional solution rather than a deep learning approach and also explores relevant questions about how to bridge the gap between learning and classic knowledge. We also expect an insightful discussion about ethical implications of traditional vision in comparison to deep learning approaches.

https://sites.google.com/view/tradicv/home

ECCV
EUROPEAN CONFERENCE ON COMPUTER VISION
MILANO 2024

max planck institut
informatik

Scientific Papers *vs* Literary Fiction

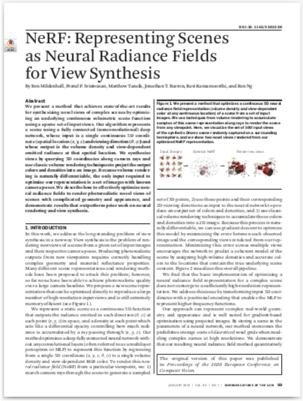


William Shakespeare

Der Sturm / The Tempest

Zweisprachige Ausgabe (Deutsch-Englisch)
Bilingual edition (German-English)

vs.



DOI:10.1145/3503250

NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis

By Ben Mildenhall, Pratul P. Srinivasan, Matthew Tancik, Jonathan T. Barron, Ravi Ramamoorthi, and Ren Ng

Abstract

We present a method that achieves state-of-the-art results for synthesizing novel views of complex scenes by optimizing an underlying continuous volumetric scene function using a sparse set of input views. Our algorithm represents a scene using a fully connected (nonconvolutional) deep network, whose input is a single continuous 5D coordinate (spatial location (x, y, z) and viewing direction (θ, ϕ)) and whose output is the volume density and view-dependent emitted radiance at that spatial location. We synthesize views by querying 5D coordinates along camera rays and use classic volume rendering techniques to project the output colors and densities into an image. Because volume rendering is naturally differentiable, the only input required to optimize our representation is a set of images with known camera poses. We describe how to effectively optimize neural radiance fields to render photorealistic novel views of scenes with complicated geometry and appearance, and demonstrate results that outperform prior work on neural rendering and view synthesis.

Figure 1. We present a method that optimizes a continuous 5D neural radiance field representation (volume density and view-dependent color at any continuous location) of a scene from a set of input images. We use techniques from volume rendering to accumulate samples of this scene representation along rays to render the scene from any viewpoint. Here, we visualize the set of 100 input views of the synthetic *Drums* scene randomly captured on a surrounding hemisphere, and we show two novel views rendered from our optimized NeRF representation.

1. INTRODUCTION

In this work, we address the long-standing problem of view synthesis in a new way. View synthesis is the problem of rendering new views of a scene from a given set of input images and their respective camera poses. Producing photorealistic outputs from new viewpoints requires correctly handling complex geometry and material reflectance properties. Many different scene representations and rendering methods have been proposed to attack this problem; however, so far none have been able to achieve photorealistic quality over a large camera baseline. We propose a new scene representation that can be optimized directly to reproduce a large number of high-resolution input views and is still extremely memory-efficient (see Figure 1).

We represent a static scene as a continuous 5D function that outputs the radiance emitted in each direction (θ, ϕ) at each point (x, y, z) in space, and a density at each point which acts like a differential opacity controlling how much radiance is accumulated by a ray passing through (x, y, z). Our method optimizes a deep fully connected neural network without any convolutional layers (often referred to as a multilayer perceptron or MLP) to represent this function by regressing from a single 5D coordinate (x, y, z, θ, ϕ) to a single volume density and view-dependent RGB color. To render this *neural radiance field* (NeRF) from a particular viewpoint, we: 1) march camera rays through the scene to generate a sampled set of 3D points, 2) use those points and their corresponding 2D viewing directions as input to the neural network to produce an output set of colors and densities, and 3) use classical volume rendering techniques to accumulate those colors and densities into a 2D image. Because this process is naturally differentiable, we can use gradient descent to optimize this model by minimizing the error between each observed image and the corresponding views rendered from our representation. Minimizing this error across multiple views encourages the network to predict a coherent model of the scene by assigning high-volume densities and accurate colors to the locations that contain the true underlying scene content. Figure 2 visualizes this overall pipeline.

We find that the basic implementation of optimizing a neural radiance field representation for a complex scene does not converge to a sufficiently high-resolution representation. We address this issue by transforming input 5D coordinates with a positional encoding that enables the MLP to represent higher frequency functions.

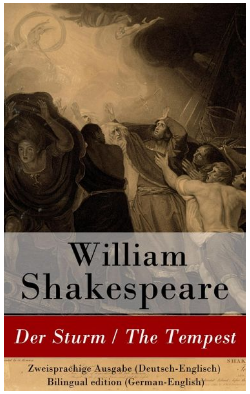
Our approach can represent complex real-world geometry and appearance and is well suited for gradient-based optimization using projected images. By storing a scene in the parameters of a neural network, our method overcomes the prohibitive storage costs of *discretized* voxel grids when modeling complex scenes at high resolutions. We demonstrate that our resulting neural radiance field method quantitatively

The original version of this paper was published in *Proceedings of the 2020 European Conference on Computer Vision*.

JANUARY 2022 | VOL. 65 | NO. 1 | COMMUNICATIONS OF THE ACM 99

Images: https://www.britannica.com/biography/William-Shakespeare/Shakespeare-the-poet-and-dramatist (left)
Mildenhall *et al.*, Communications of the ACM, 2021. (right)

Scientific Papers *vs* Literary Fiction

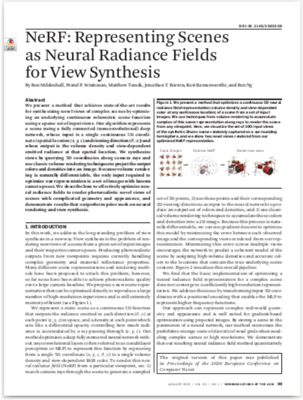


William Shakespeare

Der Sturm / The Tempest

Zweisprachige Ausgabe (Deutsch-Englisch)
Bilingual edition (German-English)

vs.



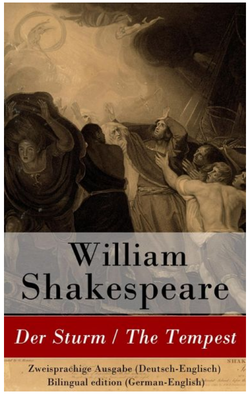
NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis

By Ben Mildenhall, Pratul P. Srinivasan, Matthew Tancik, Jonathan T. Barron, Ravi Ramamoorthi, and Ren Ng

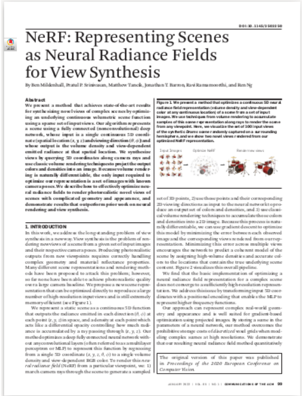
1) Tells a story
2) Covers wide range of topics
3) Written for general audience
4) Self-published (or by a publisher)
5) Less formal writing

Images: https://www.britannica.com/biography/William-Shakespeare/Shakespeare-the-poet-and-dramatist (left)
Mildenhall *et al.*, Communications of the ACM, 2021. (right)

Scientific Papers *vs* Literary Fiction


William
Shakespeare
Der Sturm / The Tempest
Zweisprachige Ausgabe (Deutsch-Englisch)
Bilingual edition (German-English)

vs.



DOI:10.1145/3503250

NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis

By Ben Mildenhall, Pratul P. Srinivasan, Matthew Tancik, Jonathan T. Barron, Ravi Ramamoorthi, and Ren Ng

Abstract

We present a method that achieves state-of-the-art results for synthesizing novel views of complex scenes by optimizing an underlying continuous volumetric scene function using a sparse set of input views. Our algorithm represents a scene using a fully connected (nonconvolutional) deep network, whose input is a single continuous 5D coordinate (spatial location (x, y, z) and viewing direction (θ, ϕ)) and whose output is the volume density and view-dependent emitted radiance at that spatial location. We synthesize views by querying 5D coordinates along camera rays and use classic volume rendering techniques to project the output colors and densities into an image. Because volume rendering is naturally differentiable, the only input required to optimize our representation is a set of images with known camera poses. We describe how to effectively optimize neural radiance fields to render photorealistic novel views of scenes with complicated geometry and appearance, and demonstrate results that outperform prior work on neural rendering and view synthesis.

1. INTRODUCTION

In this work, we address the long-standing problem of view synthesis in a new way. View synthesis is the problem of rendering new views of a scene from a given set of input images and their respective camera poses. Producing photorealistic outputs from new viewpoints requires correctly handling complex geometry and material reflectance properties. Many different scene representations and rendering methods have been proposed to attack this problem; however, so far none have been able to achieve photorealistic quality over a large camera baseline. We propose a new scene representation that can be optimized directly to reproduce a large number of high-resolution input views and is still extremely memory-efficient (see Figure 1).

We represent a static scene as a continuous 5D function that outputs the radiance emitted in each direction (θ, ϕ) at each point (x, y, z) in space, and a density at each point which acts like a differential opacity controlling how much radiance is accumulated by a ray passing through (x, y, z). Our method optimizes a deep fully connected neural network without any convolutional layers (often referred to as a multilayer perceptron or MLP) to represent this function by regressing from a single 5D coordinate (x, y, z, θ, ϕ) to a single volume density and view-dependent RGB color. To render this *neural radiance field* (NeRF) from a particular viewpoint, we: 1) march camera rays through the scene to generate a sampled set of 3D points, 2) use those points and their corresponding 2D viewing directions as input to the neural network to produce an output set of colors and densities, and 3) use classical volume rendering techniques to accumulate those colors and densities into a 2D image. Because this process is naturally differentiable, we can use gradient descent to optimize this model by minimizing the error between each observed image and the corresponding views rendered from our representation. Minimizing this error across multiple views encourages the network to predict a coherent model of the scene by assigning high-volume densities and accurate colors to the locations that contain the true underlying scene content. Figure 2 visualizes this overall pipeline.

We find that the basic implementation of optimizing a neural radiance field representation for a complex scene does not converge to a sufficiently high-resolution representation. We address this issue by transforming input 5D coordinates with a positional encoding that enables the MLP to represent higher frequency functions.

Our approach can represent complex real-world geometry and appearance and is well suited for gradient-based optimization using projected images. By storing a scene in the parameters of a neural network, our method overcomes the prohibitive storage costs of *discretized* voxel grids when modeling complex scenes at high resolutions. We demonstrate that our resulting neural radiance field method quantitatively

Figure 1. We present a method that optimizes a continuous 5D neural radiance field representation (volume density and view-dependent color at any continuous location) of a scene from a set of input images. We use techniques from volume rendering to accumulate samples of this scene representation along rays to render the scene from any viewpoint. Here, we visualize the set of 100 input views of the synthetic *Drums* scene randomly captured on a surrounding hemisphere, and we show two novel views rendered from our optimized NeRF representation.

The original version of this paper was published in *Proceedings of the 2020 European Conference on Computer Vision*.

JANUARY 2022 | VOL. 65 | NO. 1 | COMMUNICATIONS OF THE ACM 99

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

Images: https://www.britannica.com/biography/William-Shakespeare/Shakespeare-the-poet-and-dramatist (left)
Mildenhall *et al.*, Communications of the ACM, 2021. (right)

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

$$X = \left\{ \begin{array}{l} \text{new ideas, theories, algorithms, neural architectures} \\ \text{solutions to existing (e.g., long-standing) and new problems} \\ \text{combinations of components (existing and new)} \\ \text{current state of the art} \\ \text{opinion on a certain topic} \end{array} \right\}$$

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

IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE

ACMSIGGRAPH

CVPR JUNE 17-21, 2024 SEATTLE, WA

ECCV

More types?

Paper and Publication Types

Full paper

Journal Article

Conference Paper (Proceedings)

Workshop Paper

quality

predicted impact

nature

IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE

ACMSIGGRAPH

CVPR JUNE 17-21, 2024 SEATTLE, WA

ECCV

More types?

Short paper

Survey/STAR

Opinion

Corrigendum

Technical Report (e.g., on arXiv)

Dissertation

Book

Textbook

Academic Writing

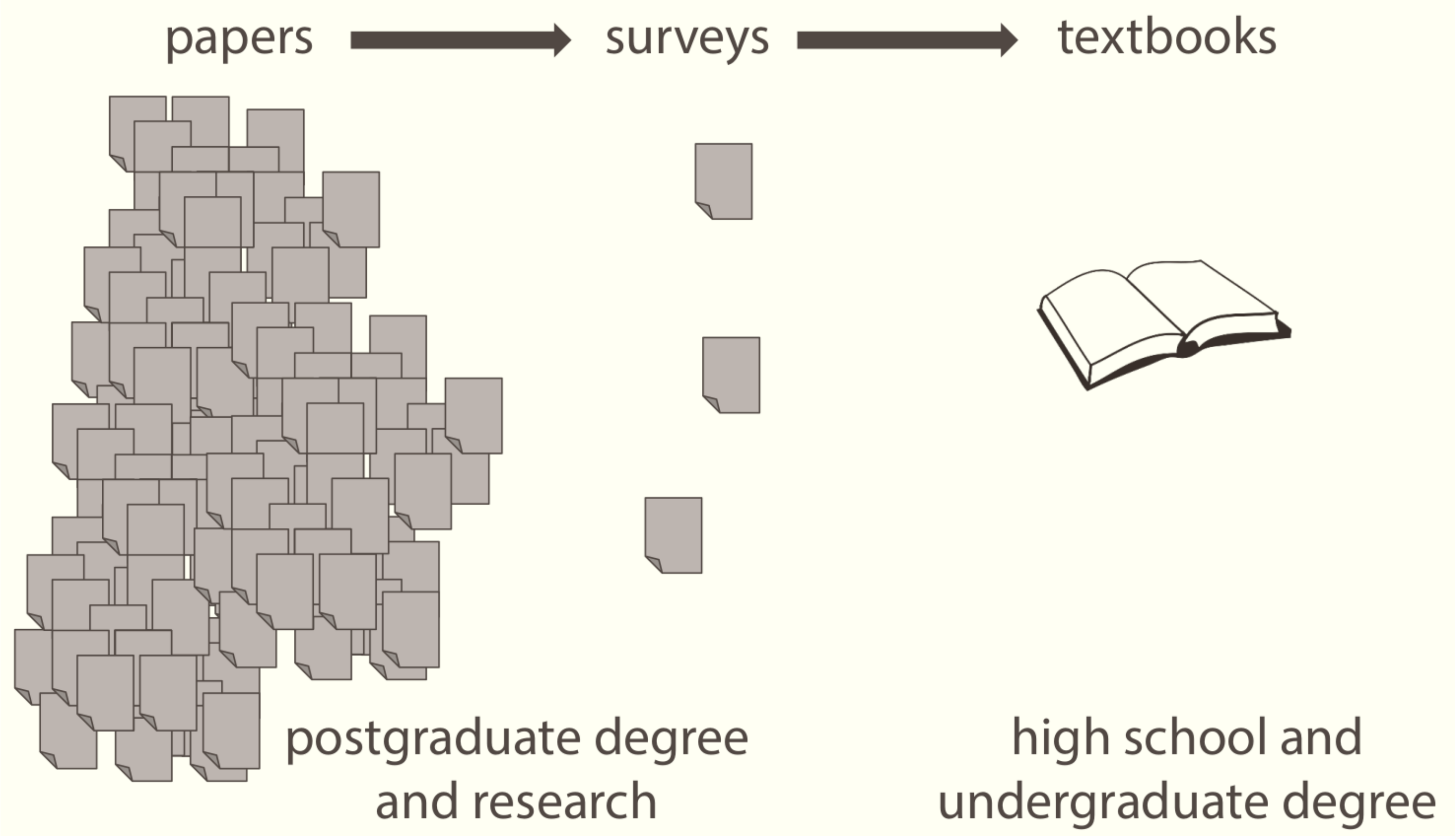

papers
surveys
textbooks
postgraduate degree and research
high school and undergraduate degree

Image by Weipeng Xu

Research and Papers

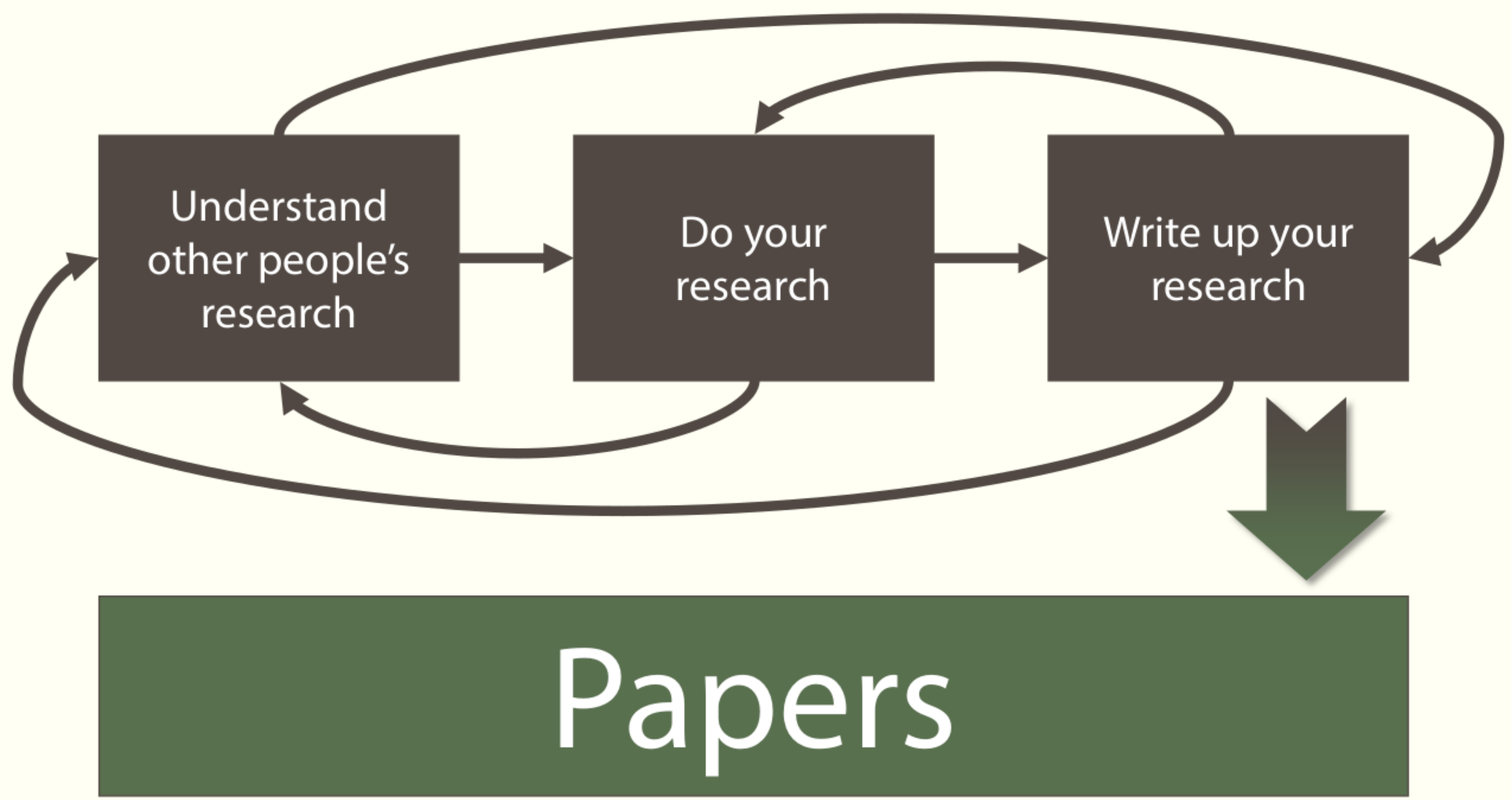
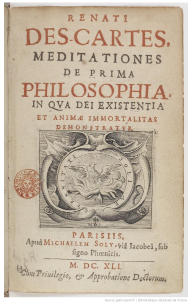


Image by Weipeng Xu

The Reviewing Process



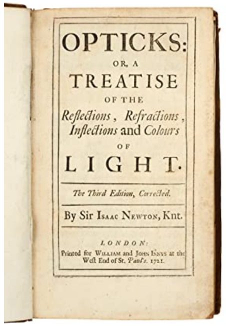
RENATI
DES-CARTES,
MEDITATIONES
DE PRIMA
PHILOSOPHIA,
IN QVA DEI EXISTENTIA
ET ANIMÆ IMMORTALITAS
DEMONSTRATVR.

AETERNITATI SOLI
MXS

PARISIIS,
Apud MICHAELEM SOLY, viâ Iacobeâ, ſub
ſigno Phœnicis.

M. DC. XLI.
Cum Priuilegio, & Approbatione Doctorum.

Source gallica.bnf.fr / Bibliothèque nationale de France

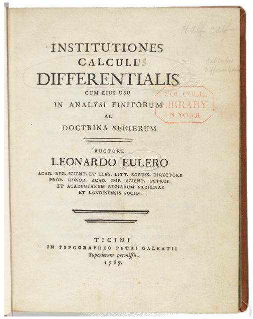


OPTICKS:
OR, A
TREATISE
OF THE
Reflections, *Refractions*,
Inflections and *Colours*
OF
LIGHT.

The Third Edition, Corrected.

By Sir ISAAC NEWTON, Knt.

LONDON:
Printed for WILLIAM and JOHN INNYS at the
Weſt End of St. *Paul's*. 1721.



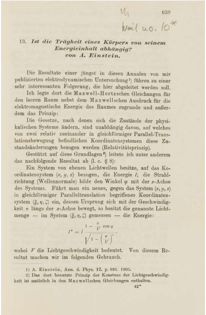
Half calf

INSTITUTIONES
CALCULI
DIFFERENTIALIS
CUM EIUS USU
IN ANALYSI FINITORUM
AC
DOCTRINA SERIERUM

Calculus
Differentialis

AUCTORE
LEONARDO EULERO
ACAD. REG. SCIENT. ET ELEG. LITT. BORUSS. DIRECTORE
PROF. HONOR. ACAD. IMP. SCIENT. PETROP.
ET ACADEMIARUM REGIARUM PARISINAE
ET LONDINENSIS SOCIO.

TICINI
IN TYPOGRAPHEO PETRI GALEATII
Superiorum permiſſu.
1787.



639

Weil no. 10*

13. ***Ist die Trägheit eines Körpers von seinem Energieinhalt abhängig?***
von A. Einstein.

Die Resultate einer jüngst in diesen Annalen von mir publizierten elektrodynamischen Untersuchung[1]) führen zu einer sehr interessanten Folgerung, die hier abgeleitet werden soll.

Ich legte dort die Maxwell-Hertzschen Gleichungen für den leeren Raum nebst dem Maxwellschen Ausdruck für die elektromagnetische Energie des Raumes zugrunde und außerdem das Prinzip:

Die Gesetze, nach denen sich die Zustände der physikalischen Systeme ändern, sind unabhängig davon, auf welches von zwei relativ zueinander in gleichförmiger Parallel-Translationsbewegung befindlichen Koordinatensystemen diese Zustandsänderungen bezogen werden (Relativitätsprinzip).

Gestützt auf diese Grundlagen[2]) leitete ich unter anderem das nachfolgende Resultat ab (l. c. § 8):

Ein System von ebenen Lichtwellen besitze, auf das Koordinatensystem (x, y, z) bezogen, die Energie l; die Strahlrichtung (Wellennormale) bilde den Winkel φ mit der x-Achse des Systems. Führt man ein neues, gegen das System (x, y, z) in gleichförmiger Paralleltranslation begriffenes Koordinatensystem (ξ, η, ζ) ein, dessen Ursprung sich mit der Geschwindigkeit v längs der x-Achse bewegt, so besitzt die genannte Lichtmenge — im System (ξ, η, ζ) gemessen — die Energie:

$$l^* = l \frac{1 - \frac{v}{V} \cos \varphi}{\sqrt{1 - \left(\frac{v}{V}\right)^2}},$$

wobei V die Lichtgeschwindigkeit bedeutet. Von diesem Resultat machen wir im folgenden Gebrauch.

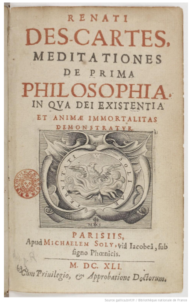
[1]) A. Einstein, Ann. d. Phys. 17. p. 891. 1905.

[2]) Das dort benutzte Prinzip der Konstanz der Lichtgeschwindigkeit ist natürlich in den Maxwellschen Gleichungen enthalten.

42*

back then

The Reviewing Process

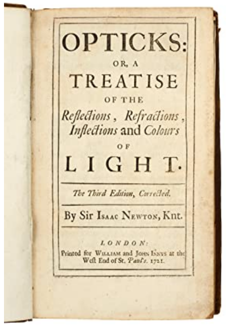


RENATI
DES-CARTES,
MEDITATIONES
DE PRIMA
PHILOSOPHIA,
IN QVA DEI EXISTENTIA
ET ANIMÆ IMMORTALITAS
DEMONSTRATVR.

PARISIIS,
Apud MICHAELEM SOLY, viâ Iacobeâ, ſub
ſigno Phœnicis.

M. DC. XLI.
Cum Priuilegio, & Approbatione Doctorum.

Source gallica.bnf.fr / Bibliothèque nationale de France

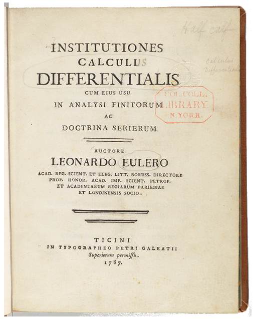


OPTICKS:
OR, A
TREATISE
OF THE
Reflections, Refractions, Inflections and *Colours*
OF
LIGHT.

The Third Edition, Corrected.

By Sir ISAAC NEWTON, Knt.

LONDON:
Printed for WILLIAM and JOHN INNYS at the
Weſt End of St. *Paul's*. 1721.



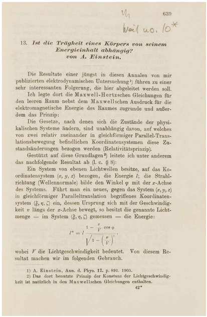
Half calf

INSTITUTIONES
CALCULI
DIFFERENTIALIS
CUM EIUS USU
IN ANALYSI FINITORUM
AC
DOCTRINA SERIERUM

calculus differentialis

AUCTORE
LEONARDO EULERO
ACAD. REG. SCIENT. ET ELEG. LITT. BORUSS. DIRECTORE
PROF. HONOR. ACAD. IMP. SCIENT. PETROP.
ET ACADEMIARUM REGIARUM PARISINAE
ET LONDINENSIS SOCIO.

TICINI
IN TYPOGRAPHEO PETRI GALEATII
Superiorum permiſſu.
1787.



U1 639

Wail no. 10*

13. ***Ist die Trägheit eines Körpers von seinem Energieinhalt abhängig?***
von A. Einstein.

Die Resultate einer jüngst in diesen Annalen von mir publizierten elektrodynamischen Untersuchung[1]) führen zu einer sehr interessanten Folgerung, die hier abgeleitet werden soll.

Ich legte dort die Maxwell-Hertzschen Gleichungen für den leeren Raum nebst dem Maxwellschen Ausdruck für die elektromagnetische Energie des Raumes zugrunde und außerdem das Prinzip:

Die Gesetze, nach denen sich die Zustände der physikalischen Systeme ändern, sind unabhängig davon, auf welches von zwei relativ zueinander in gleichförmiger Parallel-Translationsbewegung befindlichen Koordinatensystemen diese Zustandsänderungen bezogen werden (Relativitätsprinzip).

Gestützt auf diese Grundlagen[2]) leitete ich unter anderem das nachfolgende Resultat ab (l. c. § 8):

Ein System von ebenen Lichtwellen besitze, auf das Koordinatensystem (x, y, z) bezogen, die Energie l; die Strahlrichtung (Wellennormale) bilde den Winkel φ mit der x-Achse des Systems. Führt man ein neues, gegen das System (x, y, z) in gleichförmiger Paralleltranslation begriffenes Koordinatensystem (ξ, η, ζ) ein, dessen Ursprung sich mit der Geschwindigkeit v längs der x-Achse bewegt, so besitzt die genannte Lichtmenge — im System (ξ, η, ζ) gemessen — die Energie:

$$l^* = l \frac{1 - \frac{v}{V} \cos \varphi}{\sqrt{1 - \left(\frac{v}{V}\right)^2}},$$

wobei V die Lichtgeschwindigkeit bedeutet. Von diesem Resultat machen wir im folgenden Gebrauch.

[1]) A. Einstein, Ann. d. Phys. 17. p. 891. 1905.

[2]) Das dort benutzte Prinzip der Konstanz der Lichtgeschwindigkeit ist natürlich in den Maxwellschen Gleichungen enthalten.

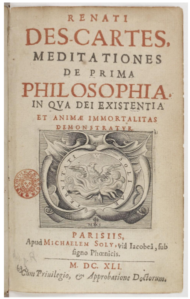
42*

back then

2000+ papers at
CVPR each year

~20k pages

The Reviewing Process

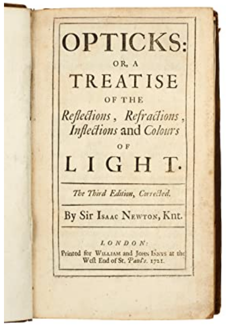


RENATI
DES-CARTES,
MEDITATIONES
DE PRIMA
PHILOSOPHIA,
IN QVA DEI EXISTENTIA
ET ANIMÆ IMMORTALITAS
DEMONSTRATVR.

PARISIIS,
Apud MICHAELEM SOLY, viâ Iacobeâ, ſub
ſigno Phœnicis.

M. DC. XLI.
Cum Priuilegio, & Approbatione Doctorum.

Source gallica.bnf.fr / Bibliothèque nationale de France

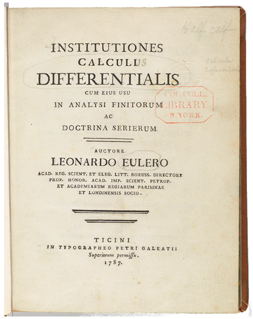


OPTICKS:
OR, A
TREATISE
OF THE
Reflections, *Refractions*,
Inflections and *Colours*
OF
LIGHT.

The Third Edition, Corrected.

By Sir ISAAC NEWTON, Knt.

LONDON:
Printed for WILLIAM and JOHN INNYS at the
Weſt End of St. *Paul's*. 1721.



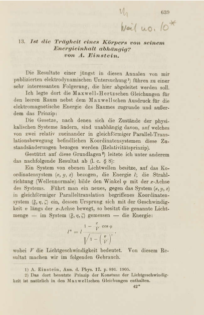
Half calf

Calculus
Differentialis

INSTITUTIONES
CALCULUS
DIFFERENTIALIS
CUM EIUS USU
IN ANALYSI FINITORUM
AC
DOCTRINA SERIERUM

AUCTORE
LEONARDO EULERO
ACAD. REG. SCIENT. ET ELEG. LITT. BORUSS. DIRECTORE
PROF. HONOR. ACAD. IMP. SCIENT. PETROP.
ET ACADEMIARUM REGIARUM PARISINAE
ET LONDINENSIS SOCIO.

TICINI
IN TYPOGRAPHEO PETRI GALEATII
Superiorum permiſſu.
1787.



Weil no. 10*

639

13. ***Ist die Trägheit eines Körpers von seinem Energieinhalt abhängig?* von A. Einstein.**

Die Resultate einer jüngst in diesen Annalen von mir publizierten elektrodynamischen Untersuchung[1]) führen zu einer sehr interessanten Folgerung, die hier abgeleitet werden soll.

Ich legte dort die Maxwell-Hertzschen Gleichungen für den leeren Raum nebst dem Maxwellschen Ausdruck für die elektromagnetische Energie des Raumes zugrunde und außerdem das Prinzip:

Die Gesetze, nach denen sich die Zustände der physikalischen Systeme ändern, sind unabhängig davon, auf welches von zwei relativ zueinander in gleichförmiger Parallel-Translationsbewegung befindlichen Koordinatensystemen diese Zustandsänderungen bezogen werden (Relativitätsprinzip).

Gestützt auf diese Grundlagen[2]) leitete ich unter anderem das nachfolgende Resultat ab (l. c. § 8):

Ein System von ebenen Lichtwellen besitze, auf das Koordinatensystem (x, y, z) bezogen, die Energie l; die Strahlrichtung (Wellennormale) bilde den Winkel φ mit der x-Achse des Systems. Führt man ein neues, gegen das System (x, y, z) in gleichförmiger Paralleltranslation begriffenes Koordinatensystem (ξ, η, ζ) ein, dessen Ursprung sich mit der Geschwindigkeit v längs der x-Achse bewegt, so besitzt die genannte Lichtmenge — im System (ξ, η, ζ) gemessen — die Energie:

$$l^* = l \frac{1 - \frac{v}{V} \cos \varphi}{\sqrt{1 - \left(\frac{v}{V}\right)^2}},$$

wobei V die Lichtgeschwindigkeit bedeutet. Von diesem Resultat machen wir im folgenden Gebrauch.

[1]) A. Einstein, Ann. d. Phys. 17. p. 891. 1905.

[2]) Das dort benutzte Prinzip der Konstanz der Lichtgeschwindigkeit ist natürlich in den Maxwellschen Gleichungen enthalten.

42*

back then

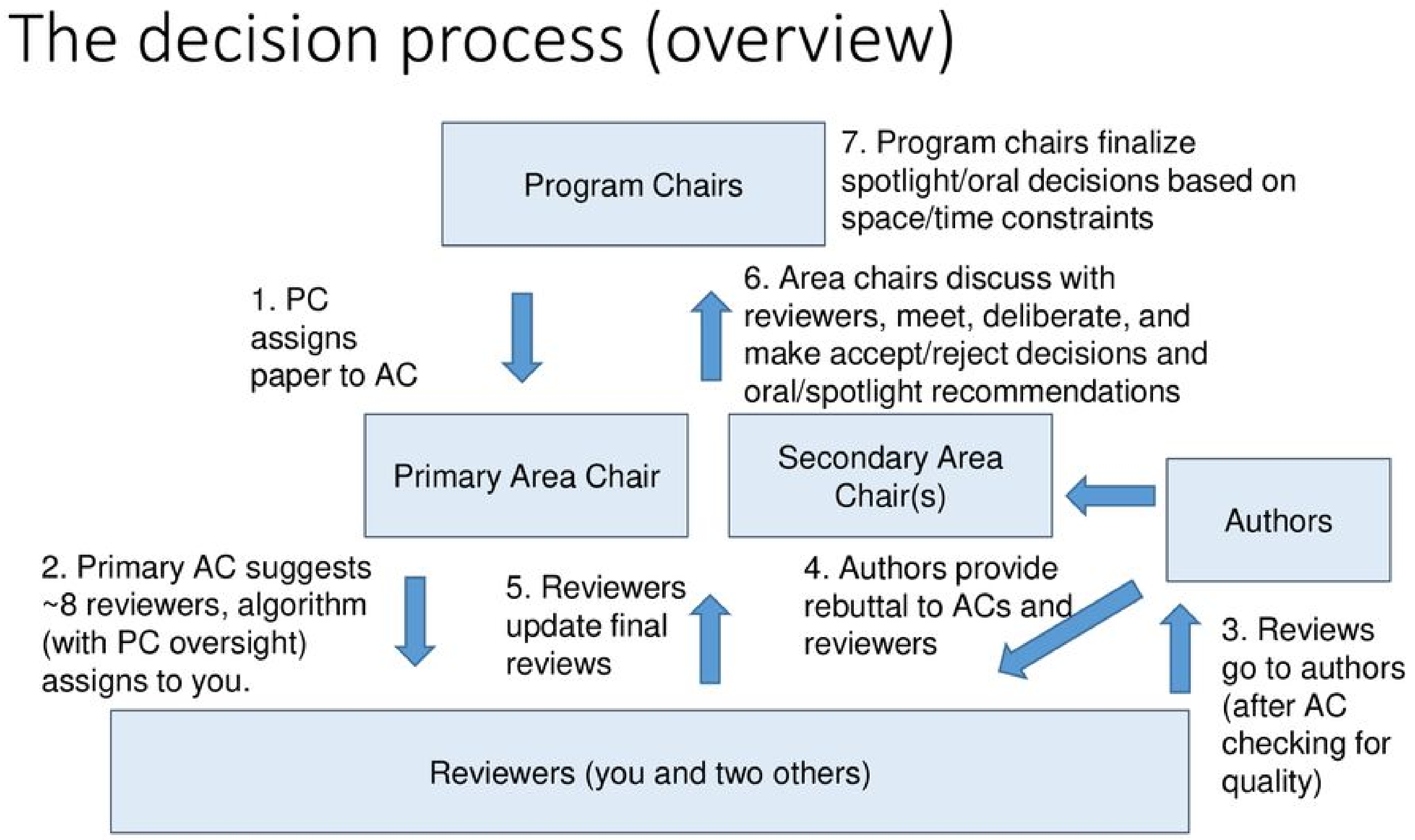
2000+ papers at
CVPR each year

~20k pages

+

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

The Reviewing Process



The decision process (overview)

Program Chairs

7. Program chairs finalize spotlight/oral decisions based on space/time constraints

1. PC assigns paper to AC

6. Area chairs discuss with reviewers, meet, deliberate, and make accept/reject decisions and oral/spotlight recommendations

Primary Area Chair

Secondary Area Chair(s)

Authors

2. Primary AC suggests ~8 reviewers, algorithm (with PC oversight) assigns to you.

5. Reviewers update final reviews

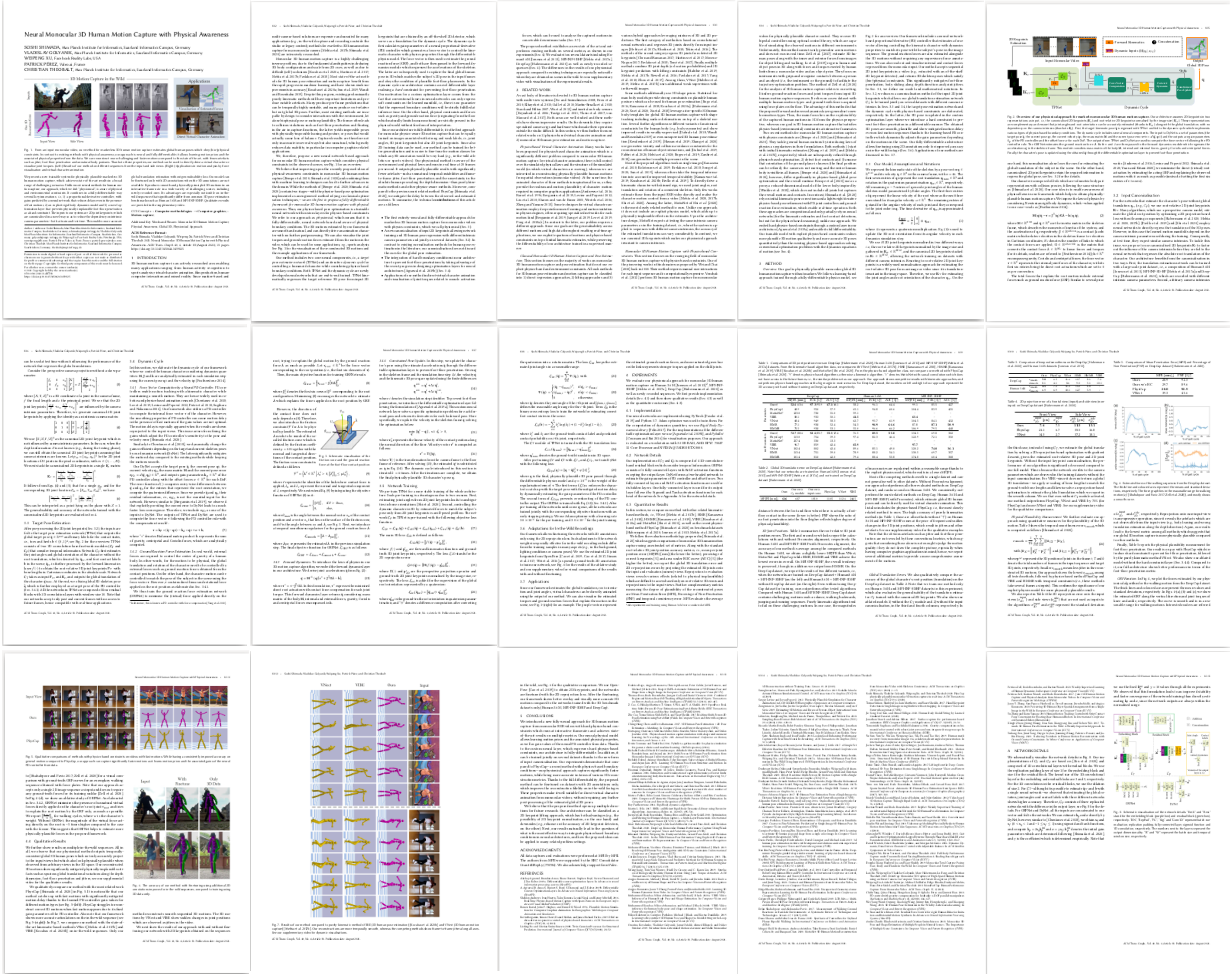
4. Authors provide rebuttal to ACs and reviewers

3. Reviews go to authors (after AC checking for quality)

Reviewers (you and two others)

Image Source: *How to write good reviews for CVPR*, by CVPR 2019 Program Chairs

Paper Structure



Shimada *et al.*, SIGGRAPH 2021.

Paper Structure

Neural Monocular 3D Human Motion Capture with Physical Awareness

max planck institut
informatik

Shimada *et al.*, SIGGRAPH 2021.

Paper Structure

Title / Header

🟥 Abstract

🟨 1. Introduction

🟧 2. Related Work

🟩 3. Method

🟦 4. Experiments

⬜ 5. Conclusions

Acknowledgements

References

Appendix

Shimada *et al.*, SIGGRAPH 2021.

Paper Structure

- Title / Header
- 🟥 Abstract
- 🟨 1. Introduction
- 🟧 2. Related Work
- 🟩 3. Method
- 🟦 4. Experiments
- ⬜ 5. Conclusions
- Acknowledgements
- References
- Appendix

Video Poster

Webpage

Source Code

SIGGRAPH 2021

Neural Monocular
3D Human Motion Capture with Physical Awareness

Soshi Shimada[1,2], Vladislav Golyanik[1,2], Weipeng Xu[3], Patrick Pérez[4], Christian Theobalt[1,2]

[1] Max Planck Institute for Informatics, Visual Computing and Artificial Intelligence Department
[2] Saarland Informatics Campus,
[3] facebook Reality Labs,
[4] Valeo.ai

Shimada *et al.*, SIGGRAPH 2021.

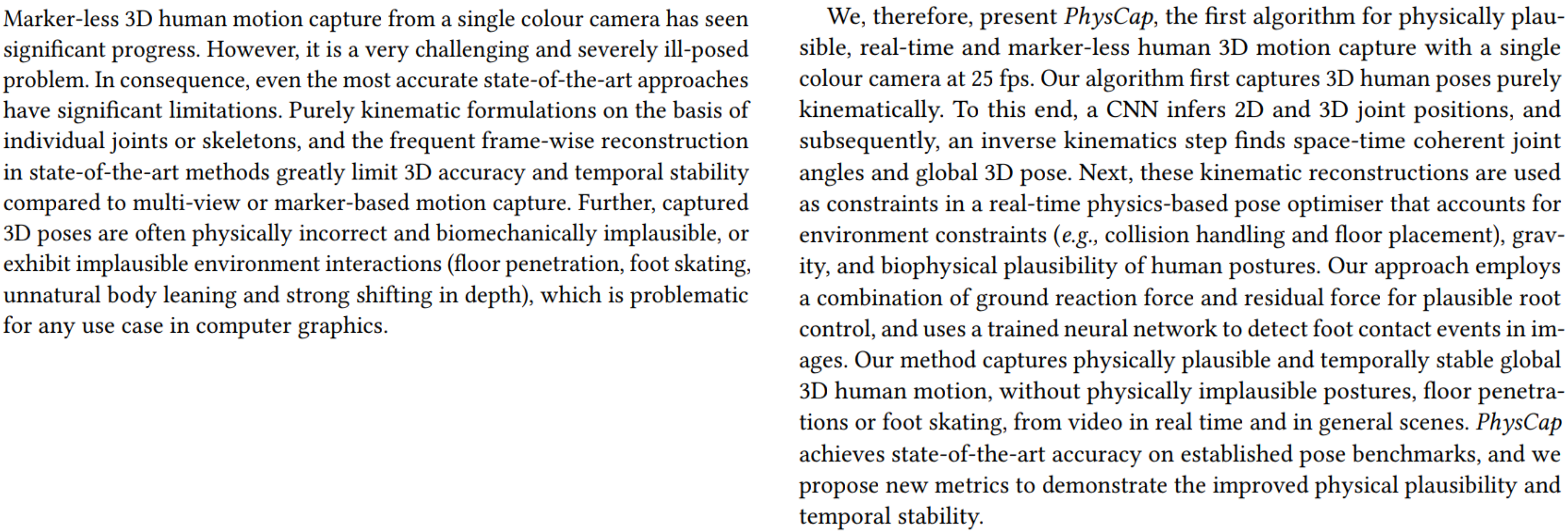
Paper Structure

- Abstract
- 1. Introduction
- 2. Related Work
- 3, 4, 5. Method
- 6. Results
  - 6.1 Datasets
  - 6.2 Comparisons
  - 6.3 Discussion
- 7. Conclusion
- Acknowledgements

NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis

Mildenhall *et al.*, Communications of the ACM, 2021.

Abstract

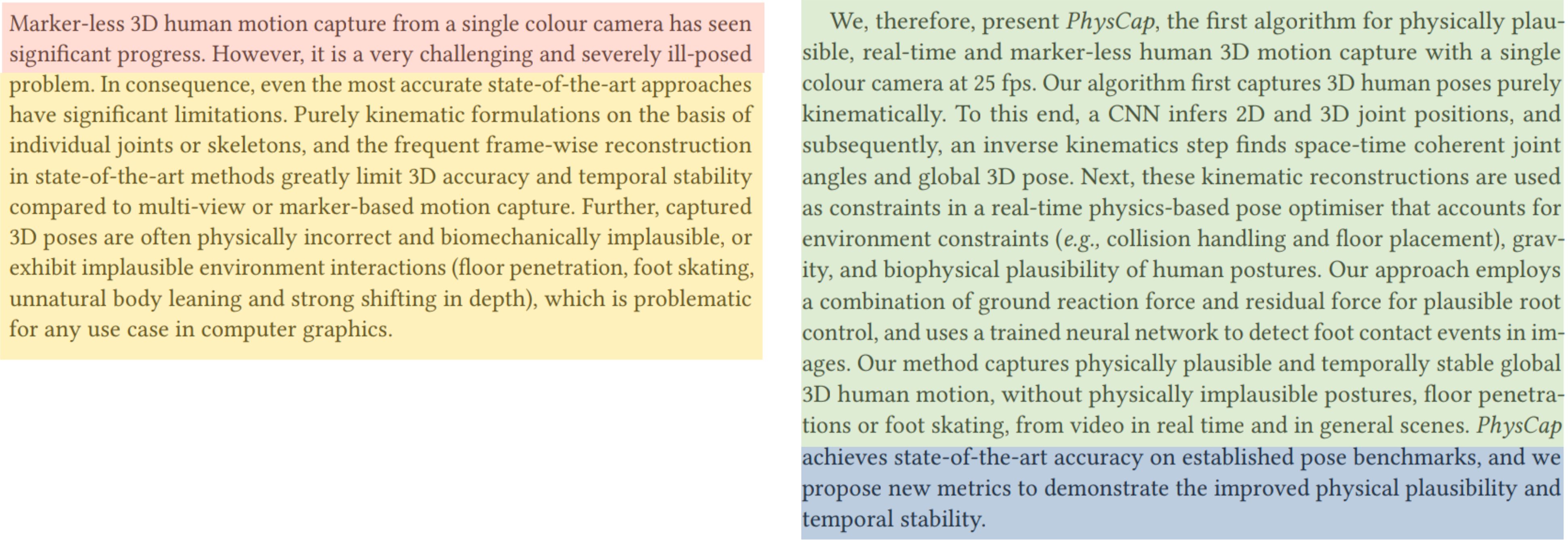


Marker-less 3D human motion capture from a single colour camera has seen significant progress. However, it is a very challenging and severely ill-posed problem. In consequence, even the most accurate state-of-the-art approaches have significant limitations. Purely kinematic formulations on the basis 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.

Shimada *et al.*, SIGGRAPH ASIA 2020.

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

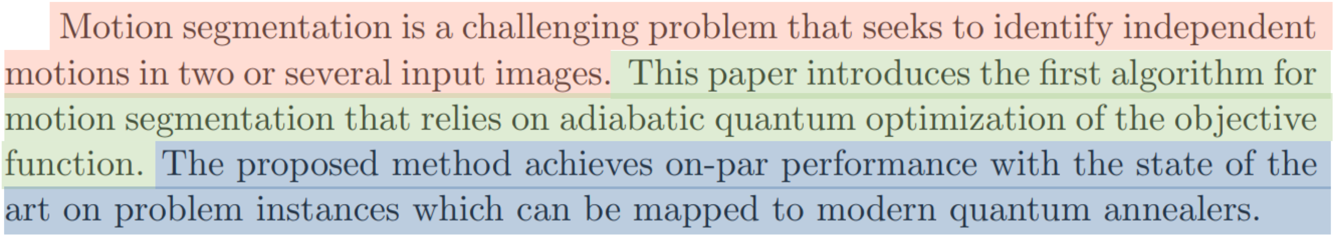
🟩 method/contributions

🟨 challenges

🟦 experimental set-up and results

Shimada *et al.*, SIGGRAPH ASIA 2020.

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

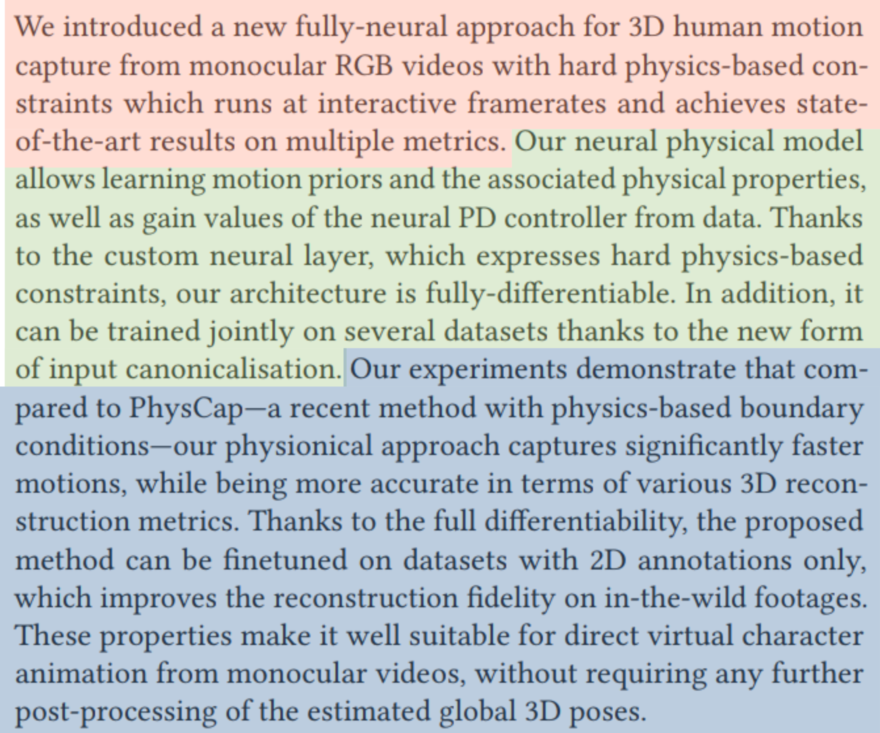
🟩 method/contributions

🟨 challenges

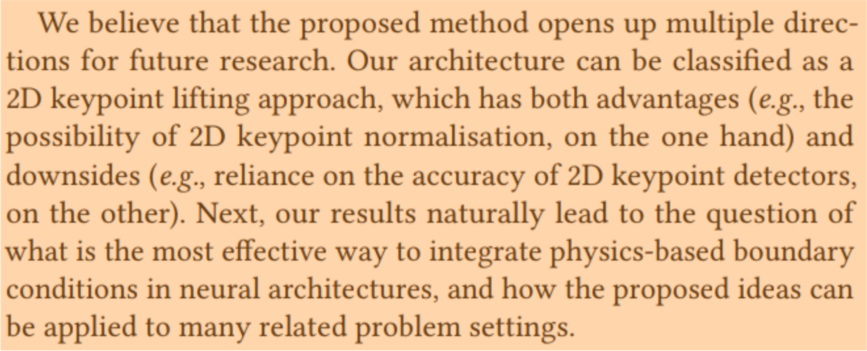
🟦 experimental set-up and results

Arrigoni *et al.*, ECCV 2022.

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.

🟥 problem

🟩 method/contributions

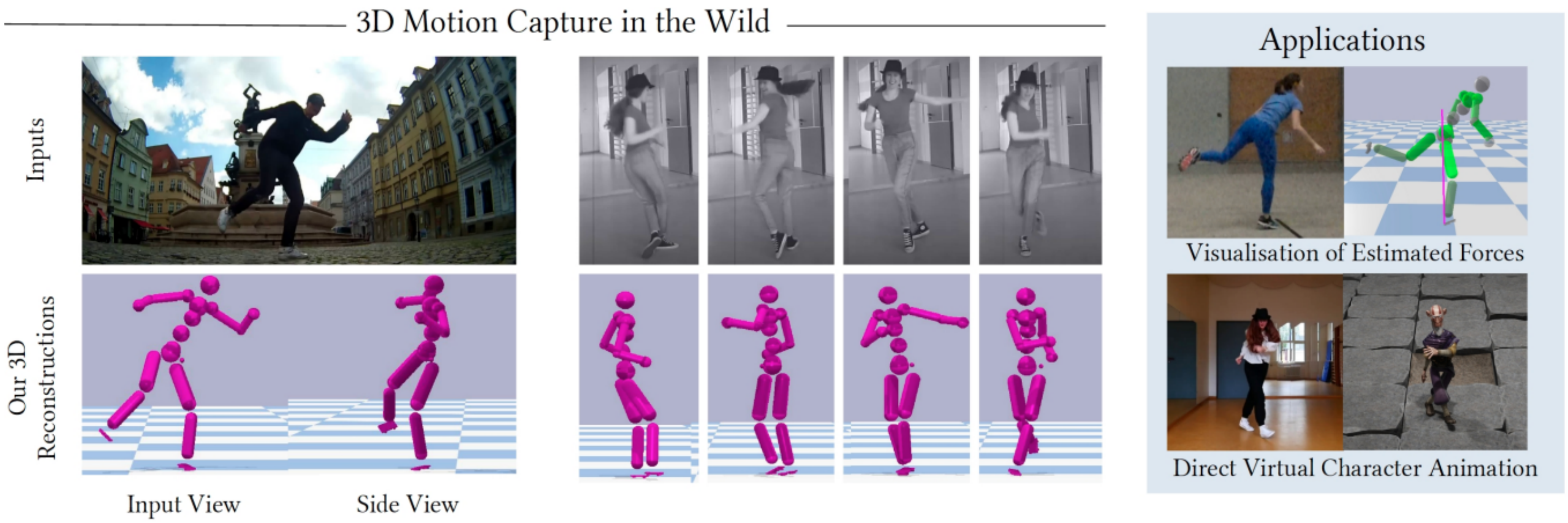
🟨 challenges

🟦 experimental results

🟧 outlook

Shimada *et al.*, SIGGRAPH ASIA 2020.

Introduction

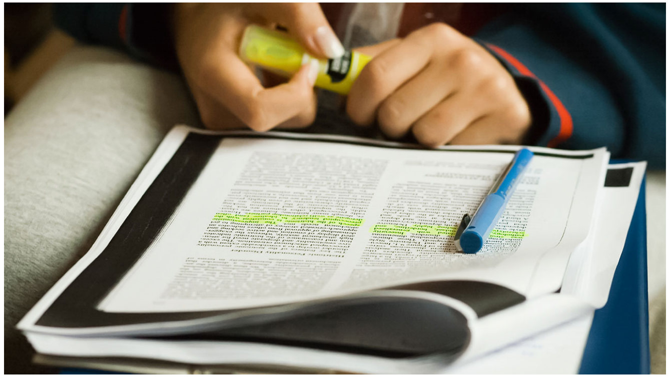

3D Motion Capture in the Wild
Inputs
Our 3D Reconstructions
Input View
Side View
Applications
Visualisation of Estimated Forces
Direct Virtual Character Animation

- **What** is the problem?
- **Why** is it important and difficult?
- Drawbacks of previous approaches
- **How?** Key aspects of the proposed approach
- A paper often explicitly states technical contributions

- Possible locations for contributions:
  - **Towards the end of the Intro section**
  - At the end of the Related Work section
  - At the end of the Overview of the Method section

Shimada *et al.*, SIGGRAPH 2021.

Related Work



- Purpose: The authors **showcase their expertise** in the field.
- One the one hand: RW is not a history of the field and not a survey
- On the other hand: RW is **not a simple enumeration!**
- Possible structure: Centred around the criteria of the main contributions
- Categorises and classifies works; **a “cone-like” structure is common.**
- Often discusses works related to the proposed approach and states 1) which ones are most closely related and 2) **in which aspects the paper at hand differs from them.**

Images: https://www.science.org/content/article/how-seriously-read-scientific-paper
https://pngtree.com/freebackground/pile-of-papers-on-a-table_3180731.html

Method Section

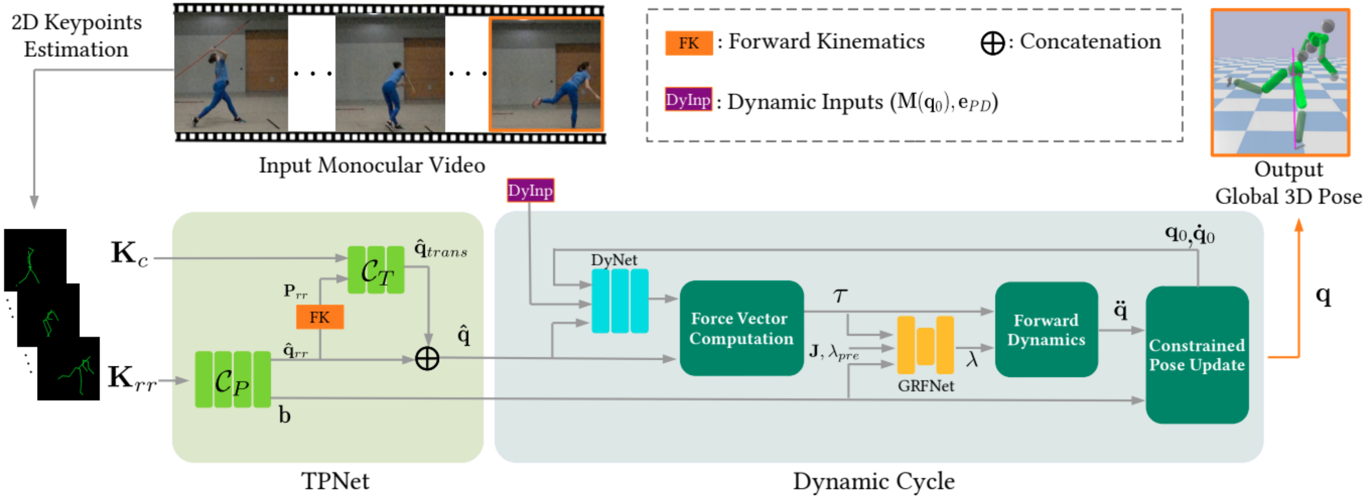


Fig. 2. **Overview of our physionical approach for markerless monocular 3D human motion capture.** Our architecture assumes 2D keypoints in two representations as input, *i.e.*, the canonicalised 2D keypoints ($\mathbf{K}_c$) and root-relative 2D keypoints normalised by the image size ($\mathbf{K}_{rr}$). These representations are complementary and ensure that joint angles and root orientation can be accurately estimated (thanks to $\mathbf{K}_{rr}$) along with the global translation, with no dependency on the camera intrinsics (thanks to $\mathbf{K}_c$). First, the target kinematic pose $\hat{\mathbf{q}}$ is regressed with TPNet and fed to the dynamic cycle which implements various types of physics-based boundary conditions. The dynamic cycle includes several neural components. The input to DyNet is a set of parameters (the target pose $\hat{\mathbf{q}}$, the current pose $\mathbf{q}_0$, the current velocity $\dot{\mathbf{q}}_0$, the mass matrix $\mathbf{M}$ and the current pose error $\mathbf{e}_{\mathrm{PD}} = d(\hat{\mathbf{q}}, \mathbf{q}_0)$) and the outputs are gain parameters k_p of the PD controller and the offset force α for each DoF. The outputs from TPNet and DyNet are used to compute the force vector $\boldsymbol{\tau}$ following the PD controller rule. The GRFNet estimates the ground reaction force $\boldsymbol{\lambda}$. Both $\boldsymbol{\tau}$ and λ are then passed to the forward dynamics module which regresses the accelerations $\ddot{\mathbf{q}}$ in the skeleton frame. This module considers mass matrix of the body $\mathbf{M}$, internal and external forces, gravity, Coriolis and centripetal forces. Finally, the character's pose is updated using the obtained $\ddot{\mathbf{q}}$ through the differentiable optimisation layer to prevent foot-floor penetration.

3.4.3 Forward Dynamics. To introduce the laws of physics in our 3D motion capture algorithm, we embed the forward dynamics layer in our architecture. We derive joint accelerations $\ddot{\mathbf{q}}$ from Eq. (2):

$$\ddot{\mathbf{q}} = \mathbf{M}^{-1}(\mathbf{q})(\boldsymbol{\tau}^* + \mathbf{J}^{\mathsf{T}}\mathbf{G}\boldsymbol{\lambda} - \mathbf{h}), \qquad (13)$$

where $\boldsymbol{\tau}^* = \boldsymbol{\tau} - \mathbf{J}^{\mathsf{T}}\mathbf{G}\boldsymbol{\lambda}$. In this formulation, $\boldsymbol{\tau}^*$ expresses the minimised direct root actuation with contact force compensation for each joint torque. This forward dynamics layer returns $\ddot{\mathbf{q}}$ considering mass matrix of the body $\mathbf{M}$, internal and external forces, gravity, Coriolis and centripetal forces encompassed in $\mathbf{h}$.

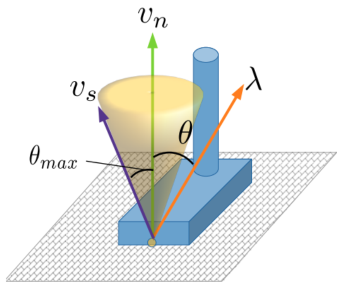
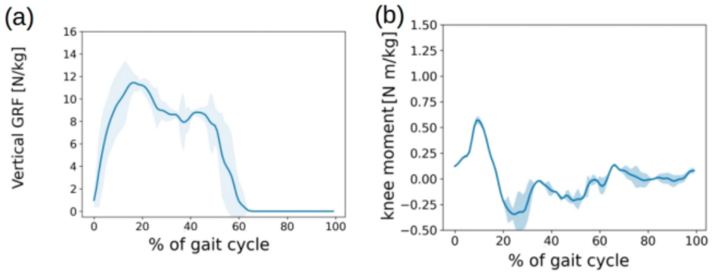


Fig. 3. Schematic visualisation of the friction cone and the ground reaction force at the foot-floor contact position.

Elements: Shimada *et al.*, SIGGRAPH 2021.

Experimental Section



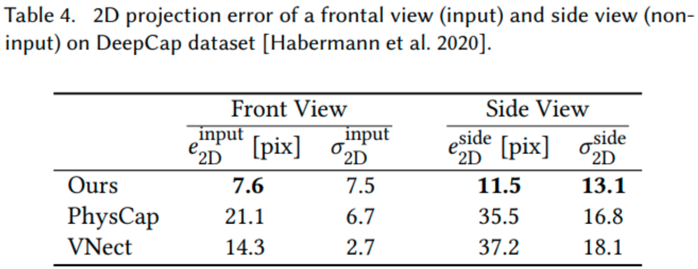


Table 4. 2D projection error of a frontal view (input) and side view (non-input) on DeepCap dataset [Habermann et al. 2020].

	Front View		Side View	
	e_{2D}^{input} [pix]	σ_{2D}^{input}	e_{2D}^{side} [pix]	σ_{2D}^{side}
Ours	**7.6**	7.5	**11.5**	**13.1**
PhysCap	21.1	6.7	35.5	16.8
VNect	14.3	2.7	37.2	18.1

Shimada *et al.*, SIGGRAPH 2021.

Tables, plots, qualitative visualisations…

- Evaluation methodology/performance metrics
- Datasets used for the evaluation
- Implementation details and processing time

- Experimental results **and their interpretation**
- **Probably Discussion (different from interpretation!)**

Scientific Writing

Main principles:

Objectivity
Precision
Clarity
Efficiency

Source: https://crk.umn.edu/sites/crk.umn.edu/files/science-writing.pdf

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**

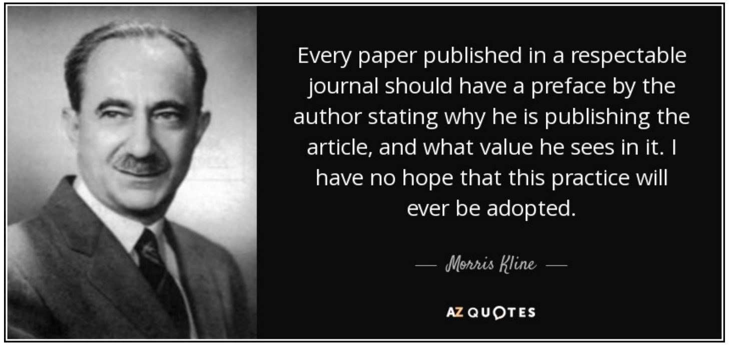
Source: https://crk.umn.edu/sites/crk.umn.edu/files/science-writing.pdf

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

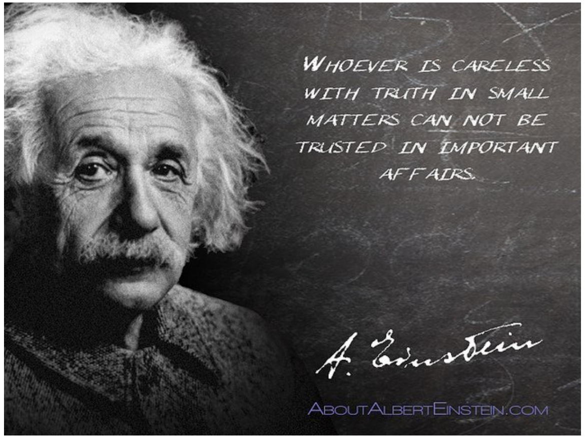
- What is the paper trying to convey?
- Why are the research and the obtained results significant?
- How were the results evaluated/measured?
- What were the results?
- What is the conclusion, and why?
- Do I trust the findings?



Every paper published in a respectable journal should have a preface by the author stating why he is publishing the article, and what value he sees in it. I have no hope that this practice will ever be adopted.

— Morris Kline —

AZ QUOTES



WHOEVER IS CARELESS WITH TRUTH IN SMALL MATTERS CAN NOT BE TRUSTED IN IMPORTANT AFFAIRS.

A. Einstein

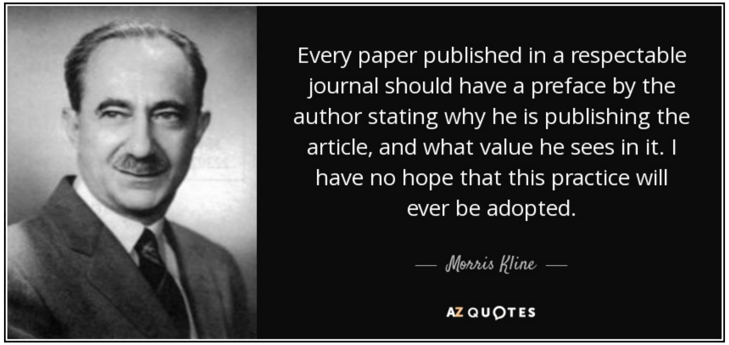
ABOUTALBERTEINSTEIN.COM

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!

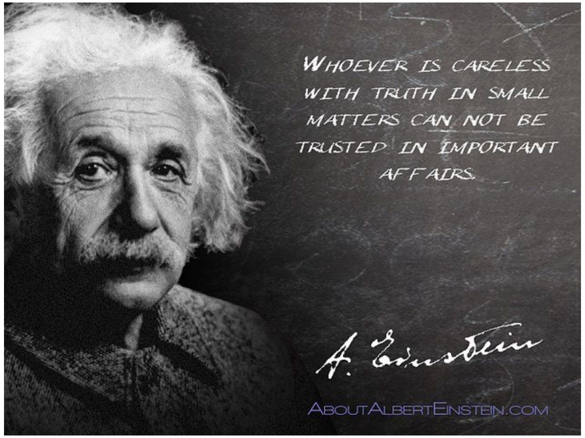
Ask questions!



Every paper published in a respectable journal should have a preface by the author stating why he is publishing the article, and what value he sees in it. I have no hope that this practice will ever be adopted.

— Morris Kline —

AZ QUOTES



WHOEVER IS CARELESS WITH TRUTH IN SMALL MATTERS CAN NOT BE TRUSTED IN IMPORTANT AFFAIRS.

A. Einstein

ABOUTALBERTEINSTEIN.COM

How to Read a Technical Paper

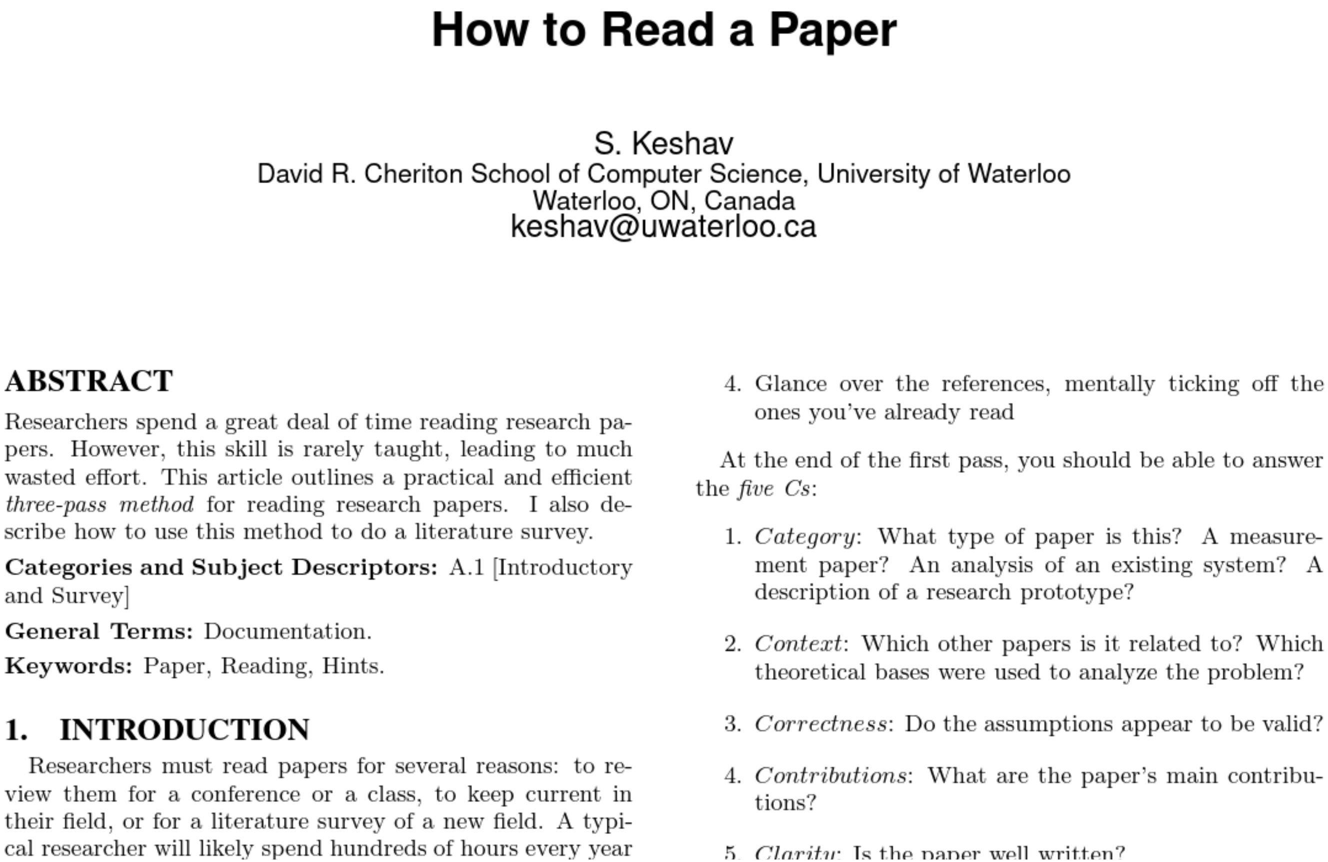
Q: What is your goal (when to stop)?

How to Read a Technical Paper

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

→ To see qualitative results	skim trough 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

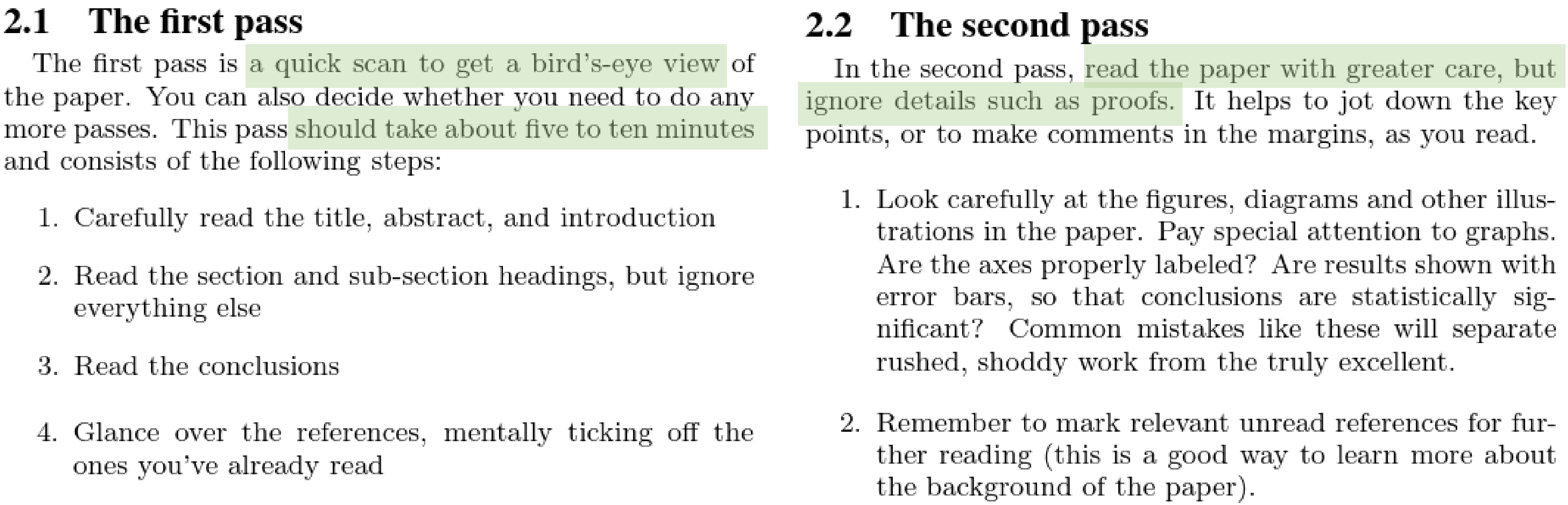
4. Glance over the references, mentally ticking off the ones you've already read

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?

Source: http://ccr.sigcomm.org/online/files/p83-keshavA.pdf

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

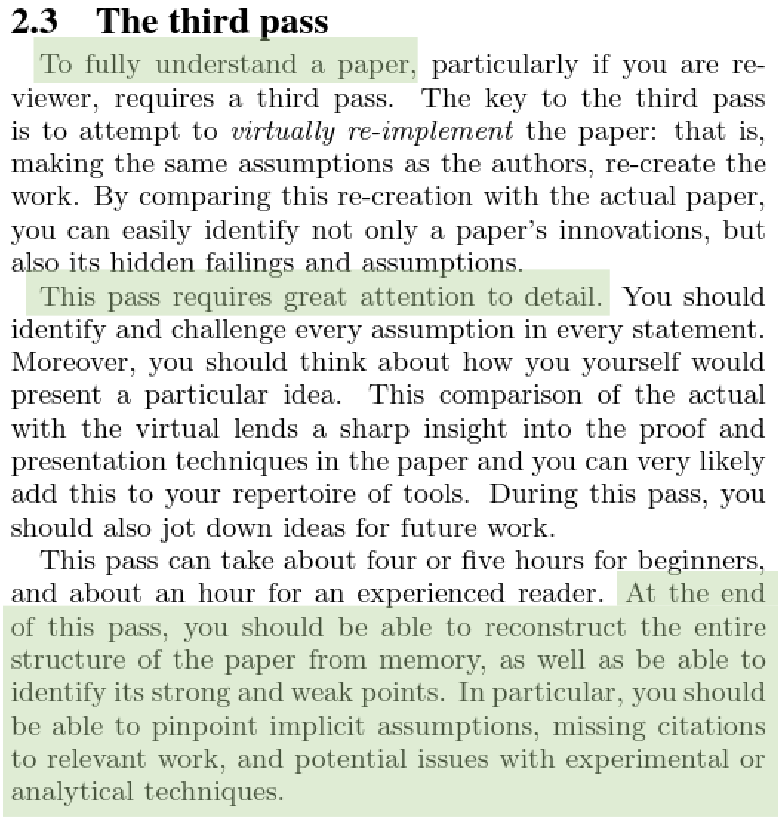
2.2 The second pass

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).

Source: http://ccr.sigcomm.org/online/files/p83-keshavA.pdf

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

Source: http://ccr.sigcomm.org/online/files/p83-keshavA.pdf

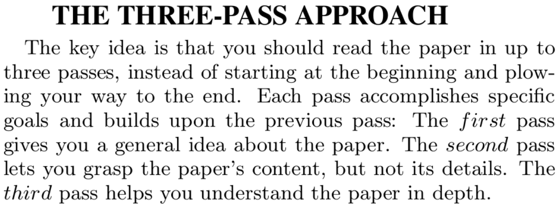
Remember What You Read

- 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?

Objectivity

Precision

Clarity

Efficiency