Deep Convolutional Neural Networks Discriminate Between Different Types of Material Kinematics

Hossein Mehrzadfar[1], Alexandra C. Schmid[2], Arash Ashrafnejad[3], Huseyin Boyaci[2,4] Katja Doerschner[2,4]

1. Graduate School of Engineering and Science, Neuroscience Program, Bilkent University, Ankara, Turkey
2. Department of Psychology, Justus-Liebig-University, Giessen, Germany
3. Department of Electrical and Electronics Engineering, Bilkent University, Ankara, Turkey
4. Department of Psychology and Neuroscience Program, Bilkent University, Ankara, Turkey

Correspondence: Katja.Doerschner@psychol.uni-giessen.de

UMRAM

Introduction

Image motion provides a powerful source of information for recognizing material properties, frequently beyond what is available in static scenes [1, 2]. Even when surface reflectance information is absent as, for example, in sparse point-light displays, humans are able to effortlessly make judgments about various material qualities [3]. Here we introduce a novel, large set of dynamic point light stimuli (Fig. 1) in order to investigate into which classes human observers and deep neural networks categorize deforming substances. Our ultimate goal is to understand which information the visual system uses in order to categorize dynamic materials.

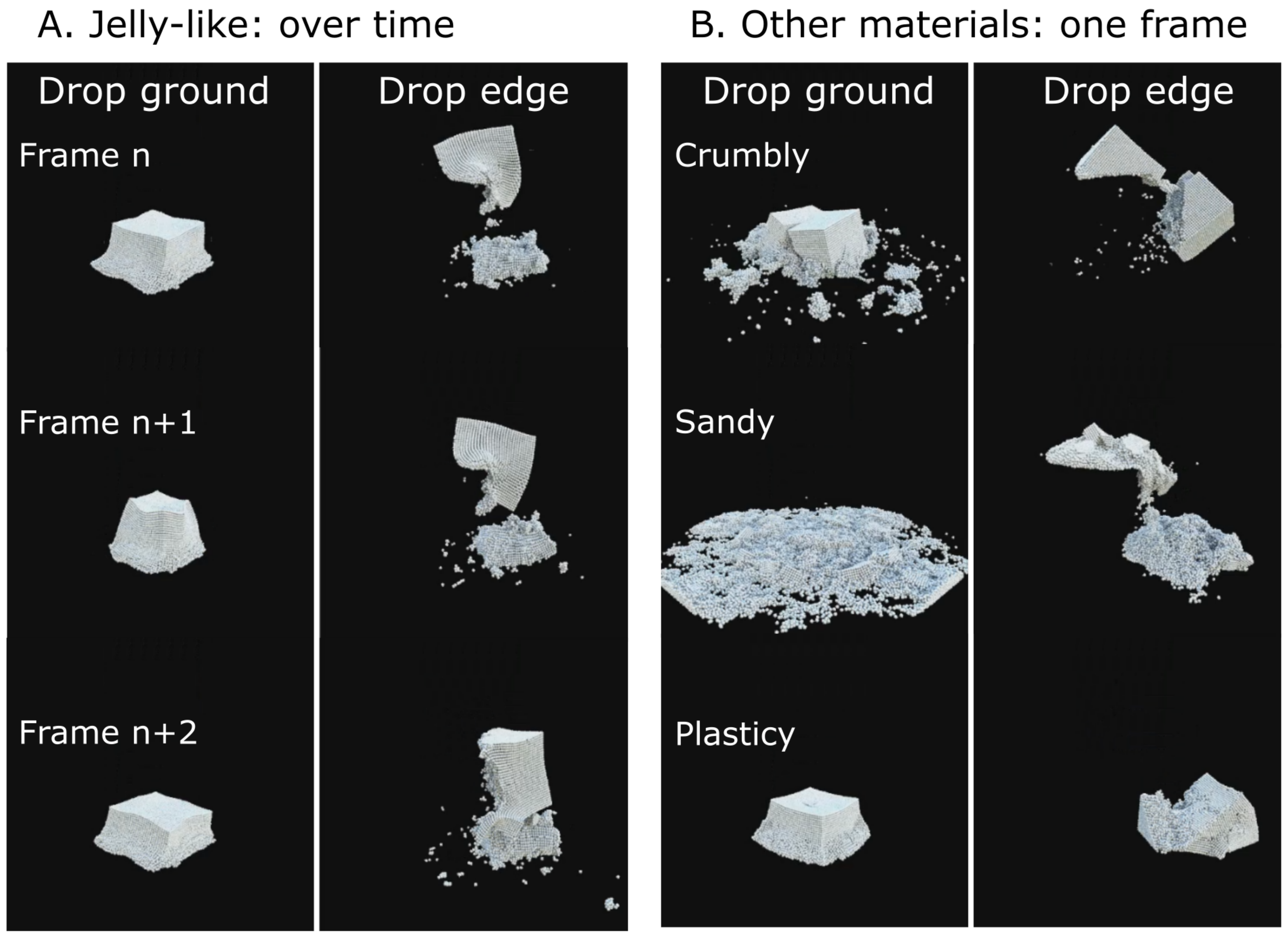

Figure 1

Stimuli differed in terms of material kinematics (stiffness, damping, collision damping, breakage), the force applied (dropping, throwing, hitting, etc.), and camera angles. Panels A shows the same jelly-like material with 2 different motions. Panel B shows 3 different materials with the same motion.

Methods

Human Categorization

In Experiment 1 (N=18) we used a subsample of stimuli, a match-to-sample paradigm and k-means clustering to establish material categories of our stimulus set (Fig. 2). In Experiment 2 (N=21) we used the same experimental paradigm - and the stimuli closest to the centroids of 8 clusters as targets - to categorize the remaining stimulus set.

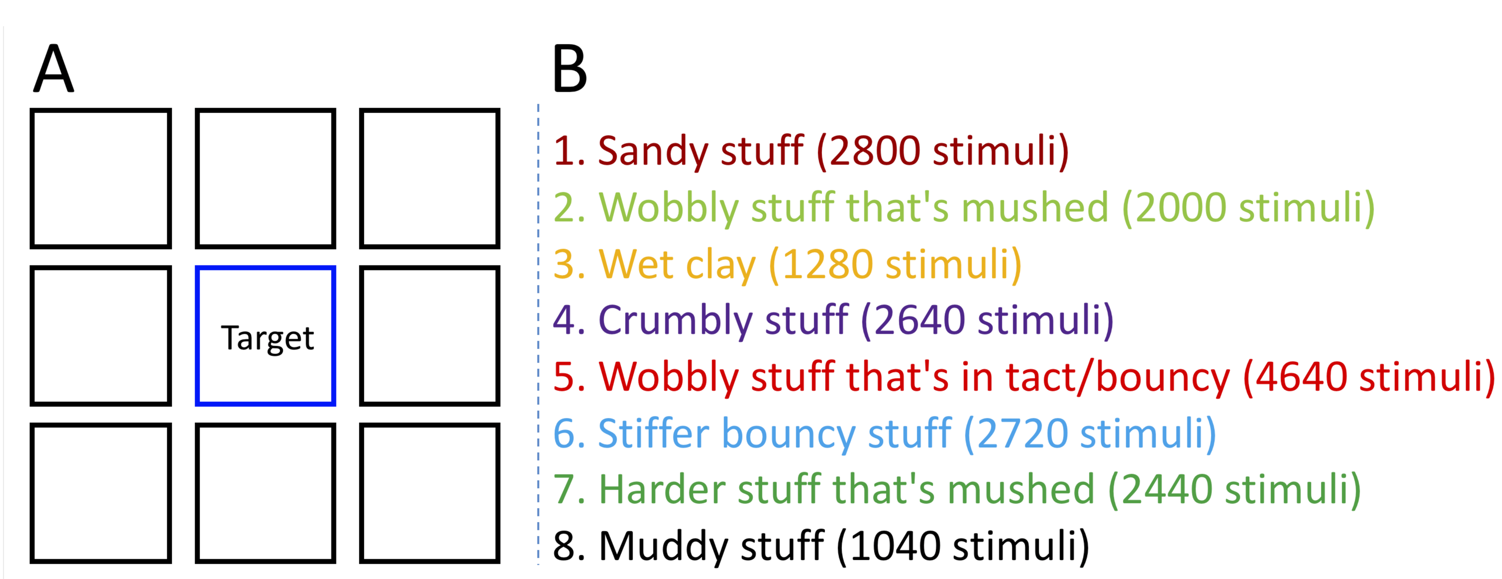

Figure 2

Panel A. On each trial observers were asked to select all stimuli that belonged to the same material category as the central target. Panel B. k-means clustering yielded about 8 categories.

Deep Convolutional Neural Networks

- Does adding a time dimension [4] substantially improve classification of dynamic materials using deep convolutional neural networks (CNN)?
- How does the performance of 3D-CNNs vary if the categorization is based on human perception versus stimulus creation parameters?

2D-CNN

- Two convolutional layers (with ReLu activation functions) each followed by a 2*2 Max-Pooling layer and two fully-connected layers at the end (Fig.3)
- Input to CNN: 1 frame (frame 17, 10th frame after the impact) of each movie. (trained with: 144 images per category)

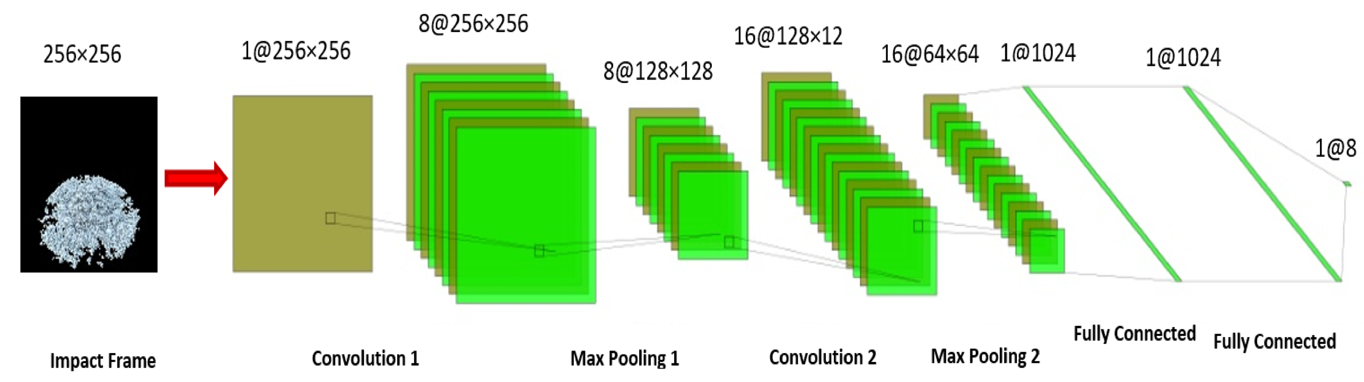

Figure 3

2D-Convolutional Neural Network architecture.

3D-CNN

- Two 3D-convolutional layers (with ReLu activation functions) each followed by a 2*2*2 3D-Max-Pooling layer and a fully-connected layer at the end (Fig.4)
- Input to CNN: 10 frames (frames 7-16, 10 frame after the impact) of each animation. (trained with: 576 videos per category for the data labeled based on rendering parameters and 408 videos per category for the data labeled based on human perception)

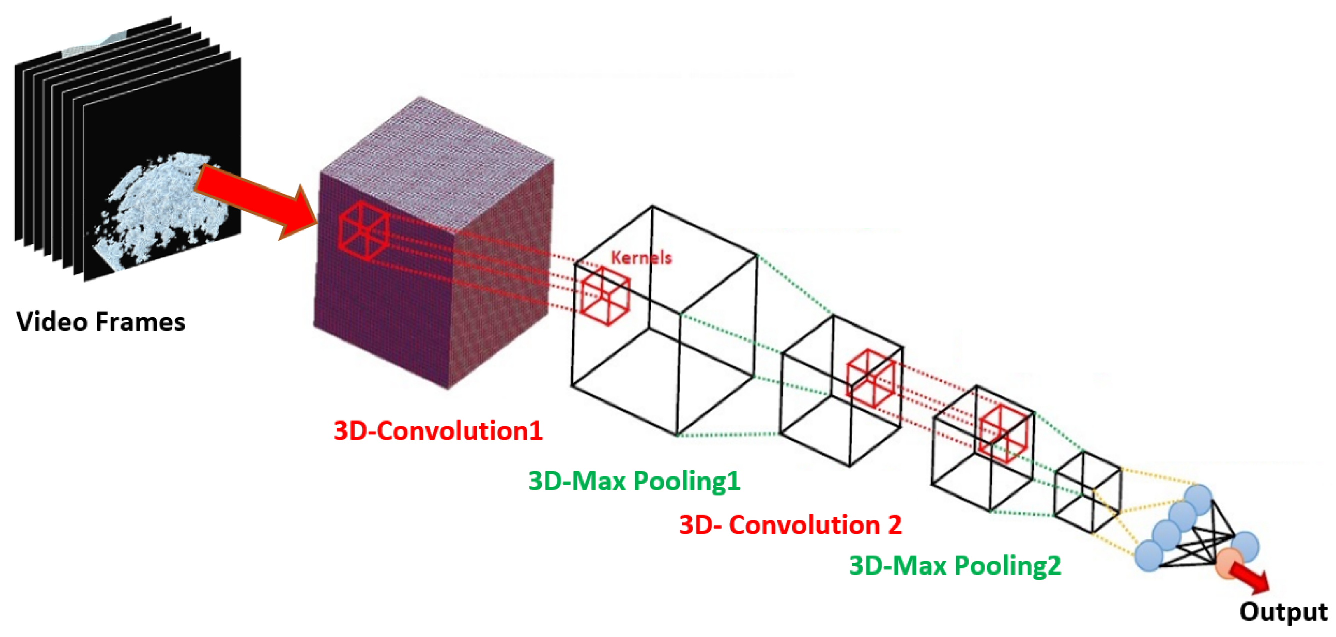

Figure 4

3D-Convolutional Neural Network architecture. Image from [5]

Results

- 2D-CNN accuracy 87.3 % (data labeled based on rendering parameters)
- 3D-CNN accuracy $\begin{cases} \text{97.8 \% (data labeled based on rendering parameters)} \\ \text{80.0 \% (data labeled based on \textbf{human perception})} \end{cases}$

Summary

1. Adding a time dimension increases the classification accuracy
2. 3D-CNN performs better if training data labeling is based on stimulus creation parameters than based on human perceptual categories.

Open Questions and Future Work

Why do physics-labels work better than human perception? What information is used? Do humans and CNNs make similar mistakes, while categorizing dynamic materials?

Acknowledgments

Funding for this project is provided by the Alexander von Humboldt foundation through a Sofja Kovalevskaja Award.

References

[1] K. Doerschner, D. Zang, D. Kersten, and P. Schrater, "Cooperative computation of shape and material from motion," *Journal of Vision*, vol. 9, pp. 51–51, mar 2010.

[2] K. Doerschner, R. W. Fleming, O. Yilmaz, P. R. Schrater, B. Hartung, and D. Kersten, "Visual Motion and the Perception of Surface Material," *Current Biology*, vol. 21, pp. 2010–2016, dec 2011.

[3] A. C. Schmid and K. Doerschner, "Shatter and splatter: The contribution of mechanical and optical properties to the perception of soft and hard breaking materials," *Journal of Vision*, vol. 18, p. 14, jan 2018.

[4] D. Tran, L. Bourdev, R. Fergus, L. Torresani, and M. Paluri, "Learning Spatiotemporal Features with 3D Convolutional Networks," tech. rep.

[5] A. Balu, S. Ghadai, K. G. Lore, G. Young, A. Krishnamurthy, and S. Sarkar, "Learning Localized Geometric Features Using 3D-CNN: An Application to Manufacturability Analysis of Drilled Holes," dec 2016.