



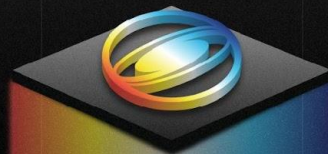
**SIGGRAPH 2024**

DENVER+ 28 JUL — 1 AUG

THE PREMIER CONFERENCE  
& EXHIBITION ON  
COMPUTER GRAPHICS &  
INTERACTIVE TECHNIQUES

# (Tiny) Problems with Web Scraping the ACM Digital Library

SIGGRAPH 2024 Village: History: The ACM  
SIGGRAPH History Archive—Moving  
Forward by Looking Backwards





**SIGGRAPH 2024**  
DENVER+ 28 JUL — 1 AUG



**Manuel Alducin**

**Information Technology Professional**

Global Affairs Canada

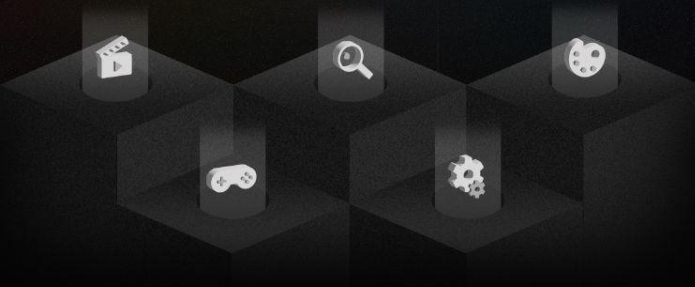
At the Embassy of Canada to Mexico

SIGGRAPH  
2024



**SIGGRAPH 2024**  
DENVER+ 28 JUL — 1 AUG

# Web Scraping



## Download

### Save some time

- Download the Publication pages
- Use filename for some configuration

## Extract Info

### Two Steps

- Get basic info from Publication page
- Get Abstract and (rebuild) References from detail page

## Write Excel

### Write File

- Except no buffering, write until the end!



- Ubuntu 22
- Python 3.10 (tried 3.11 for TOML via deadsnakes)
- Main libraries
  - BeautifulSoup: parse the HTML
  - XlsxWriter: directly write Excel file
  - requests: get detail page
- Script about 150 lines of code
- No logging (only print to screen)
- Configuration inside the script
- Script more or less worked within 2 weeks
- No Git! (what was I thinking!)

SIGGRAPH  
2024

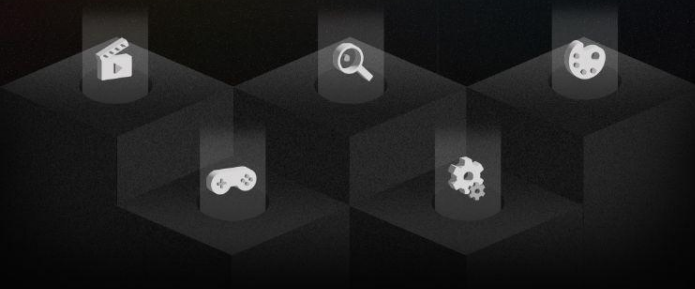


**SIGGRAPH 2024**

DENVER+ 28 JUL — 1 AUG

# The Main Problem

Time



## Late Involvement

- Learned of History Project at Vancouver 2022
- Got in touch around December 2022

## Web Scraping

- Can still require manual cleanup
- May require lots of tweaking
- Hadn't done scraping in a while

## Analyse the Problem

- Little time to really analyze the DL in-depth
- But the DL seems pretty clean, right?

## SIGGRAPH 2023 TOG

### SESSION: Geometric Optimization

RESEARCH-ARTICLE  
OPEN ACCESS

#### Winding Numbers on Discrete Surfaces

Nicole Feng, Mark Gillespie, Keenan Crane

Article No.: 36, Pages 1–17 • [https://](https://doi.org/10.1145/3592401)

[doi.org/10.1145/3592401](https://doi.org/10.1145/3592401)

In the plane, the *winding number* is the number of times a curve wraps around a given point. Winding numbers are a basic component of geometric algorithms such as point-in-polygon tests, and their generalization to data with noise or topological errors ...

2 695

```
><div class="issue-item-checkbox-container">...</div>
><div class="issue-item clearfix">
  ::before
  ><div class="issue-item_citation">...</div>
  ><div class="issue-item_content">
    ><div class="issue-item_content-right">
      ><h5 class="issue-item_title">
        <a href="/doi/10.1145/3592401">Winding Numbers of
      </h5>
      ><ul class="rlist--inline loa truncate-list trunc-d">
      ><div class="issue-item_detail">...</div>
      ><div class="issue-item_abstract truncate-text tru">
      ><div class="issue-item_footer clearfix">...</div>
    </div>
  </div>
  ::after
</div>
</div>
><div class="issue-item-container">...</div>
><div class="issue-item-container">...</div>
</div>
</div>
><div class="toc_section accordion-tabbed_tab">...</div>
><div class="toc_section accordion-tabbed_tab">...</div>
```



# The Final Output



SIGGRAPH 2024  
DENVER+ 28 JUL – 1 AUG

## SIGGRAPH 2023 Papers

learning_title	learning_conference	rel_learning_type	learning_abstract	learning_link_to_dl	learning_references	Presenters
Winding Numbers on Discrete Surfaces	SIGGRAPH 2023	Technical Papers	In the plane, the winding number is the number of times a curve	<a href="https://doi.org/10.1145/3592401">https://doi.org/10.1145/3592401</a>	65. Qingnan Zhou, Eitan Grinspun, Denis Zorin, and Alec Jacobson. 2016. Mo	Keenan Crane
Flexible Isosurface Extraction for Gradient-Based Mesh Optimizat	SIGGRAPH 2023	Technical Papers	This work considers gradient-based mesh optimization, where we	<a href="https://doi.org/10.1145/3592430">https://doi.org/10.1145/3592430</a>	74. Chenyang Zhu, Kai Xu, Siddhartha Chaudhuri, Benjiao Yi, and Hao Zhang	Jun Gao
Topology driven approximation to rational surface-surface interse	SIGGRAPH 2023	Technical Papers	Computing the intersection between two parametric surfaces (S	<a href="https://doi.org/10.1145/3592452">https://doi.org/10.1145/3592452</a>	63. Y Yamaguchi, R Kamiyama, and F Kimura. 2001. Surface-Surface Interse	Ming Li
A Practical Wave Optics Reflection Model for Hair and Fur	SIGGRAPH 2023	Technical Papers	Traditional fiber scattering models, based on ray optics, are miss	<a href="https://doi.org/10.1145/3592446">https://doi.org/10.1145/3592446</a>	57. Arno Zinke, Martin Rump, Tomás Lay, Andreas Weber, Anton Andriyenko	Steve Marschner
Anatomically Detailed Simulation of Human Torso	SIGGRAPH 2023	Technical Papers	Many existing digital human models approximate the human ske	<a href="https://doi.org/10.1145/3592425">https://doi.org/10.1145/3592425</a>	59. Jānis Savļovskis and Kristaps Raits, [n. d.]. Biomechanics of the Spine. In	C. Karen Liu
HACK: Learning a Parametric Head and Neck Model for High-fide	SIGGRAPH 2023	Technical Papers	Significant advancements have been made in developing param	<a href="https://doi.org/10.1145/3592093">https://doi.org/10.1145/3592093</a>	115. Yiyu Zhuang, Hao Zhu, Xusen Sun, and Xun Cao. 2022. Mofanet: Morph	Jingyi Yu
GestureDiffuCLIP: Gesture Diffusion Model with CLIP Latents	SIGGRAPH 2023	Technical Papers	The automatic generation of stylized co-speech gestures has rec	<a href="https://doi.org/10.1145/3592097">https://doi.org/10.1145/3592097</a>	88. Yi Zhou, Connelly Barnes, Jingwan Lu, Jimei Yang, and Hao Li. 2019. On	Libin Liu
BodyFormer: Semantics-guided 3D Body Gesture Synthesis with	SIGGRAPH 2023	Technical Papers	Automatic gesture synthesis from speech is a topic that has attr	<a href="https://doi.org/10.1145/3592456">https://doi.org/10.1145/3592456</a>	51. Yukun Zhu, Ryan Kiros, Richard S. Zemel, Ruslan Salakhutdinov, Raquel	Taku Komura
Listen, Denoise, Action! Audio-Driven Motion Synthesis with Diffu	SIGGRAPH 2023	Technical Papers	Diffusion models have experienced a surge of interest as highly e	<a href="https://doi.org/10.1145/3592458">https://doi.org/10.1145/3592458</a>	133. Vikram Voleti, Alexia Jolicoeur-Martineau, and Christopher Pal. 2022. M	Gustav Eje Henter
Contact Edit: Artist Tools for Intuitive Modeling of Hand-Object In	SIGGRAPH 2023	Technical Papers	Posing high-contact interactions is challenging and time-consum	<a href="https://doi.org/10.1145/3592117">https://doi.org/10.1145/3592117</a>	43. W. Zhao, J. Zhang, J. Min, and J. Chai. 2013. Robust RealTime Physics-Ba	Nancy Pollard
Eventfulness for Interactive Video Alignment	SIGGRAPH 2023	Technical Papers	Humans are remarkably sensitive to the alignment of visual eve	<a href="https://doi.org/10.1145/3592118">https://doi.org/10.1145/3592118</a>	50. Oliver Wang, Christopher Schroers, Henning Zimmer, Markus Gross, and	Abe Davis
FactorMatte: Redefining Video Matting for Re-Composition Tasks	SIGGRAPH 2023	Technical Papers	We propose Factor Matting, an alternative formulation of the vid	<a href="https://doi.org/10.1145/3592423">https://doi.org/10.1145/3592423</a>	46. Yunke Zhang, Chi Wang, Miaomiao Cui, Peiran Ren, Xuansong Xie, Xian	Abe Davis
Computational Long Exposure Mobile Photography	SIGGRAPH 2023	Technical Papers	Long exposure photography produces stunning imagery, represen	<a href="https://doi.org/10.1145/3592124">https://doi.org/10.1145/3592124</a>	46. Timo Zindler, Jochen Schmidt, and Heinrich Niemann. 2005. Point set reg	Yael Pritch
ShapeCoder: Discovering Abstractions for Visual Programs from	SIGGRAPH 2023	Technical Papers	We introduce ShapeCoder, the first system capable of taking a d	<a href="https://doi.org/10.1145/3592416">https://doi.org/10.1145/3592416</a>	57. Fenggen Yu, Zhiqin Chen, Manyi Li, Aditya Sanghi, Hooman Shayanli, Ali	Daniel Ritchie
The Visual Language of Fabrics	SIGGRAPH 2023	Technical Papers	We introduce text2fabric, a novel dataset that links free-text des	<a href="https://doi.org/10.1145/3592391">https://doi.org/10.1145/3592391</a>	70. Xilong Zhou, Milos Hasan, Valentin Deschaintre, Paul Guerrero, Kalyan	Belen Masia
ArrangementNet: Learning Scene Arrangements for Vectorized In	SIGGRAPH 2023	Technical Papers	We present a novel vectorized indoor modeling approach that co	<a href="https://doi.org/10.1145/3592122">https://doi.org/10.1145/3592122</a>	85. Chuhang Zou, Erşin Yumer, Jimei Yang, Duygu Ceylan, and Derek Hoiem	Li Yi
Juxtaform: Interactive visual summarization for exploratory shap	SIGGRAPH 2023	Technical Papers	We present juxtaform, a novel approach to the interactive summ	<a href="https://doi.org/10.1145/3592436">https://doi.org/10.1145/3592436</a>	70. Jun-Yan Zhu, Yong Jae Lee, and Alexei A Efros. 2014. AverageExplorer. In	Karan Singh
Patternshop: Editing Point Patterns by Image Manipulation	SIGGRAPH 2023	Technical Papers	Point patterns are characterized by their density and correlation	<a href="https://doi.org/10.1145/3592418">https://doi.org/10.1145/3592418</a>	93. Yang Zhou, Zhen Zhu, Xiang Bai, Dani Lischinski, Daniel Cohen-Or, and	Gurprit Singh
VideoDoodles: Hand-Drawn Animations on Videos with Scene-Aw	SIGGRAPH 2023	Technical Papers	We present an interactive system to ease the creation of so-calle	<a href="https://doi.org/10.1145/3592413">https://doi.org/10.1145/3592413</a>	62. Zhoutong Zhang, Forrester Cole, Richard Tucker, William T Freeman, an	Adrien Bousseau
StripMaker: Perception-driven Learned Vector Sketch Consolidat	SIGGRAPH 2023	Technical Papers	Artist sketches often use multiple overdrawn strokes to depict a	<a href="https://doi.org/10.1145/3592130">https://doi.org/10.1145/3592130</a>	64. Jerry Yin, Chenxi Liu, Rebecca Lin, Nicholas Vining, Helge Rhodin, and	Alla Sheffer
Semi-supervised reference-based sketch extraction using a contr	SIGGRAPH 2023	Technical Papers	Sketches reflect the drawing style of individual artists; therefor	<a href="https://doi.org/10.1145/3592392">https://doi.org/10.1145/3592392</a>	79. Changqing Zou, Haoran Mo, Chengyong Gao, Ruofei Du, and Hongbo Fu	Junyong Noh
Split-Lohmann Multifocal Displays	SIGGRAPH 2023	Technical Papers	This work provides the design of a multifocal display that can cre	<a href="https://doi.org/10.1145/3592110">https://doi.org/10.1145/3592110</a>	42. Tao Zhan, Jianghao Xiong, Junyu Zou, and Shin-Tson Wu. 2020. Multifoca	Aswin C. Sankaranarayanan
Etendue Expansion in Holographic Near Eye Displays through Spr	SIGGRAPH 2023	Technical Papers	In this paper, we present a novel method the etendue expansion	<a href="https://doi.org/10.1145/3592441">https://doi.org/10.1145/3592441</a>	33. Wei Yuan, Li-Hua Li, Wing-Bun Lee, and Chang-Yuen Chan. 2018. Fabric	Yoonchan Jeong
Rhizomorph: The Coordinated Function of Shoots and Roots	SIGGRAPH 2023	Technical Papers	Computer graphics has dedicated a considerable amount of effor	<a href="https://doi.org/10.1145/3592145">https://doi.org/10.1145/3592145</a>	78. Y. Zhao and J. Barbic. 2013. Interactive Authoring of Simulation-ready P	Wojtek Patulicki

SIGGRAPH  
2024

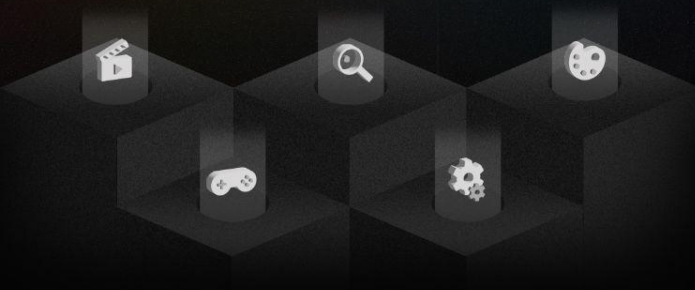


**SIGGRAPH 2024**

DENVER+ 28 JUL — 1 AUG

# Challenges

Discrepancies in the DL



# Challenges: Discrepancies in the DL

- DL is rate limited: find out how low you can go
- Updates to the DL System
- CSS specificity
- Differences between TOG and older Proceedings
- Older Proceedings might include info from several types of Sessions: Panels, Sketches, etc.



# Challenges: Discrepancies in the DL

## SIGGRAPH 2023 TOG

Sections

Volume 42, Issue 4  
August 2023

← Previous Issue

Next Issue →

Issue Downloads

SESSION: Geometric Optimization

SESSION: A Material World

SESSION: Motion Recipes and Simulation

SESSION: Character Animation: Knowing What To Do With Your Hands

SESSION: Image and Video Editing

SESSION: Geometric Abstractions: Not just for Cubists

SESSION: Magical Sketching

SESSION: XR Displays and Perception: Seeing What's in Front of Your Eyes

SESSION: Procedural Modeling

SESSION: Inverse Rendering: Does Anybody Know How I Got Here?

SESSION: Fabricating Appearance

SESSION: Contours, Conformality, Coarsening, and Coordinates

SESSION: Thin and Thinner: Modeling Shells and Hair

SESSION: Full-Body XR: Beyond The Headset

SESSION: Neural Light Transport

SESSION: Pushing the Boundaries

SESSION: Cloud Rendering: Your GPU is Sometimes Else

SESSION: Diffusion for Geometry

SESSION: Character Animation

SESSION: Colorful Topics in Imaging

SESSION: Surfaces, Strips, Lights

SESSION: Most Del: Fast, Large and Learned Deformables

SESSION: Material Rendering

SESSION: Surface Reconstruction

SESSION: Neural Capturing

SESSION: Fabrication-Chained Design

SESSION: All About Meshes

SESSION: Geometric Optimization

RESEARCH-ARTICLE

OPEN ACCESS

Winding Numbers on Discrete Surfaces

Black Fren, Mark Gilbert, Simon Gane

Article No: 36, Pages 1-17 • <https://doi.org/10.1145/3592409>

In the plane, the winding number is the number of times a curve wraps around a given point. Winding numbers are a basic component of geometric algorithms such as point-in-polygon tests, and their generalization to data with noise or topological errors ...

19 0 595

RESEARCH-ARTICLE

OPEN ACCESS

Flexible Iosurface Extraction for Gradient-Based Mesh Optimization

Tianchang Shen, Jacob Munkberg, Jan Hertzberg, Kexue Yin, Zan Wang, >5

Article No: 37, Pages 1-18 • <https://doi.org/10.1145/3592410>

This work considers gradient-based mesh optimization, where we iteratively optimize for a 3D surface mesh by representing it as the isosurface of a scalar field, an increasingly common paradigm in applications including photogrammetry, generative ...

19 0 2,495

RESEARCH-ARTICLE

OPEN ACCESS

Topology driven approximation to rational surface-surface intersection via interval algebraic topology analysis

Bo-Sun Cheng, Jingpei Zhang, Yihui Xie, Ming Li

Article No: 38, Pages 1-18 • <https://doi.org/10.1145/3592413>

Computing the intersection between two parametric surfaces (PS) is one of the most fundamental problems in geometric and solid modeling. Maintaining the 5D topology is critical to its computation robustness. We propose a topology-driven hybrid symbolic ...

19 0 517

SESSION: A Material World

SESSION: Motion Recipes and Simulation

SESSION: Character Animation: Knowing What To Do With Your Hands

SESSION: Image and Video Editing

SESSION: Geometric Abstractions: Not just for Cubists

SESSION: Magical Sketching

SESSION: XR Displays and Perception: Seeing What's in Front of Your Eyes

SESSION: Procedural Modeling

SESSION: Inverse Rendering: Does Anybody Know How I Got Here?

SESSION: Fabricating Appearance

SESSION: Contours, Conformality, Coarsening, and Coordinates

SESSION: Thin and Thinner: Modeling Shells and Hair

SESSION: Full-Body XR: Beyond The Headset

## SIGGRAPH 1995 Proceedings

ARTICLE  
FREE

Computer graphics achievement award

Kurt Akeley

Page 10 • <https://doi.org/10.1145/218380.218386>

99 0 417

ARTICLE  
FREE

Steven A. Coons award for outstanding creative contributions to computer graphics

Jose Luis Encarnação

Pages 11-12 • <https://doi.org/10.1145/218380.218388>

99 0 356

ARTICLE  
FREE

Geometry compression

Michael Deering

Pages 13-20 • <https://doi.org/10.1145/218380.218391>

99 431 2,996

ARTICLE  
FREE

Polygon-assisted JPEG and MPEG compression of synthetic images

Marc Levoy

Pages 21-28 • <https://doi.org/10.1145/218380.218392>

99 58 1,198




# Challenges: Discrepancies in the DL



## SIGGRAPH 2009 Papers



☐ **SESSION: Fast image processing and retargeting**

☐ **ARTICLE** [Session details: Fast image processing and retargeting](#)





 [Aron Hertzmann](#)

<https://doi.org/10.1145/3257573>

  0




 





☐ **RESEARCH-ARTICLE** [Gaussian KD-trees for fast high-dimensional filtering](#)

 [Andrew Adams](#),  [Natasha Gelfand](#),  [Jennifer Dolson](#),  [Marc Levoy](#)


Article No.: 21, Pages 1–12 • <https://doi.org/10.1145/1576246.1531327>

We propose a method for accelerating a broad class of non-linear filters that includes the bilateral, non-local means, and other related filters. These filters can all be expressed in a similar way: First, assign each value to be filtered a position in ...

 70  2,705 




  •  





☐ **RESEARCH-ARTICLE** [Edge-avoiding wavelets and their applications](#)

 [Raanan Fattal](#)

Article No.: 22, Pages 1–10 • <https://doi.org/10.1145/1576246.1531328>

We propose a new family of second-generation wavelets constructed using a robust data-prediction lifting scheme. The support of these new wavelets is constructed based on the edge content of the image and avoids having pixels from both sides of an edge. ...

 36  2,625 

  •  

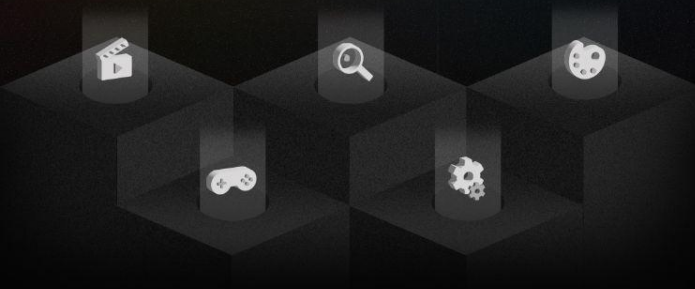
SIGGRAPH  
2024



**SIGGRAPH 2024**  
DENVER+ 28 JUL — 1 AUG

# Challenges

Dynamic Content



# Challenges: Dynamic Content

- Used for large lists of Articles and for large number of authors
- Also different between older and newer content
- Two options: try to automate or take care manually
- Automate: use Selenium?
- Manually: just click away

# Challenges: Dynamic Content

## SIGGRAPH Asia 2023

☐ SESSION: Shells

☐ RESEARCH-ARTICLE  
**C-Shells: Deployable Gridshells with Curved Beams**  
Quentin Becker, Seichi Suzuki, Yinying Ren, Davide Pellis, Julian Panetta  
Article No.: 173, Pages 1–17 • <https://doi.org/10.1145/3618356>  
We introduce a computational pipeline for simulating and designing C-shells, a new class of planar-to-spatial deployable linkage structures. A C-shell is composed of curved flexible beams connected at rotational joints that can be assembled in a stress-free state.  
192 views

☐ RESEARCH-ARTICLE  
**Kirchhoff-Love Shells with Arbitrary Hyperelastic Materials**  
Jihao Wen, Jeroen Baerle  
Article No.: 174, Pages 1–15 • <https://doi.org/10.1145/3618405>  
Kirchhoff-Love shells are commonly used in many branches of engineering, including in computer graphics, but have so far been simulated only under limited nonlinear material options. We derive the Kirchhoff-Love thin-shell mechanical energy for an arbitrary hyperelastic material.  
369 views

☐ RESEARCH-ARTICLE  
**Stable Discrete Bending by Analytic Eigensystem and Adaptive Orthotropic Geometric Stiffness**  
Zhendong Wang, Yixi Yang, Huanmin Wang  
Article No.: 175, Pages 1–16 • <https://doi.org/10.1145/3618372>  
In this paper, we address two limitations of dihedral angle based discrete bending (DAB) models, i.e. the indefiniteness of their energy Hessian and their vulnerability to geometry degeneracies. To tackle the indefiniteness issue, we present novel...  
126 views

☐ RESEARCH-ARTICLE  
**Progressive Shell Quasistatics for Unstructured Meshes**  
Jiay Eris Zhan, Jérôme Dutra, Yun (Raymond) Fan, Alec Jacobson  
Article No.: 176, Pages 1–17 • <https://doi.org/10.1145/3618388>  
Thin shell structures exhibit complex behaviors critical for modeling and design across wide-ranging applications. Capturing their mechanical response requires finely detailed, high-resolution meshes. Corresponding simulations for predicting equilibria...  
199 views

☐ SESSION: Character and Rigid Body Control

☐ SESSION: All Kinds of Meshes

☐ SESSION: Computational Design

☐ SESSION: Fluid Simulation

☐ SESSION: Robots & Characters

☐ SESSION: See Details

☐ SESSION: Motion Synthesis with Awareness

☐ SESSION: Designer Assistant

## SIGGRAPH 1999 Proceedings

☐ ARTICLE  
**The VolumePro real-time ray-casting system**  
Hanspeter Pfister, Jan Hardenbergh, Jim Knittel, Hugh Lauer, Larry Seiler  
Pages 251–260 • <https://doi.org/10.1145/311535.311563>  
207 views, 1,953 downloads

☐ ARTICLE  
**Deep compression for streaming texture intensive animations**  
Daniel Cohen-Or, Yair Mann, Shachar Fleishman  
Pages 261–267 • <https://doi.org/10.1145/311535.311564>  
37 views, 640 downloads

☐ ARTICLE  
**Optimization of mesh locality for transparent vertex caching**  
Hugues Hoppe  
Pages 269–276 • <https://doi.org/10.1145/311535.311565>  
125 views, 950 downloads

See More

Show All (+ 21)



## SIGGRAPH Asia 2023



RESEARCH-ARTICLE  
OPEN ACCESS

### SLANG.D: Fast, Modular and Differentiable Shader Programming

 [Sai Praveen Bangaru](#),  [Lifan Wu](#),  [Tzu-Mao Li](#),  [Jacob Munkberg](#),  [Gilbert Bernstein](#),  + 4

Article No.: 264, Pages 1–28 • <https://doi.org/10.1145/3618353>

We introduce SLANG.D, an extension to the Slang shading language that incorporates first-class automatic differentiation support. The new shading language allows us to transform a Direct3D-based path tracer to be fully differentiable with minor ...

 1  428 |  1



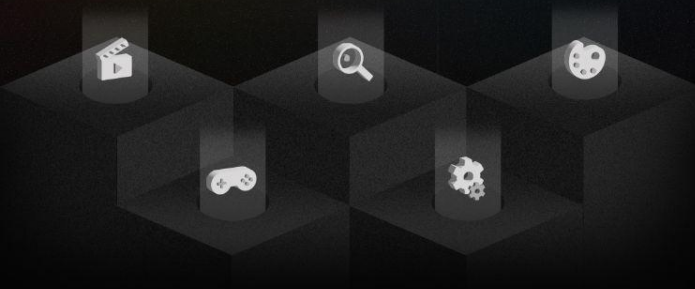
SIGGRAPH  
2024



**SIGGRAPH 2024**  
DENVER+ 28 JUL — 1 AUG

# Challenge

More Time Pressure!



# Challenge: More Time Pressure!

- Early 2023: need citations of all Papers for the Seminal Graphics Papers Vol. 2
- How do citations work in the DL?
- It's dynamic content but easy to get from page
- Re-download all Proceedings and TOG volumes again for up to date
- Re-process all HTML to get citations
- Either send the new Excels or try to merge manually

- Borked system (deadsnakes!): reinstall and switch to miniconda (Python 3.12 with TOML)
- Papers are now separated between TOG and Proceedings



SIGGRAPH  
2024



**SIGGRAPH 2024**  
DENVER+ 28 JUL — 1 AUG

**Thanks!**

