CLEVR-ER

David Nukrai, Yotam Erel
2022

What will we cover?

- Motivation

- Relationships

- Blender Implementation

- Baseline Results

- Discussion & Future Work

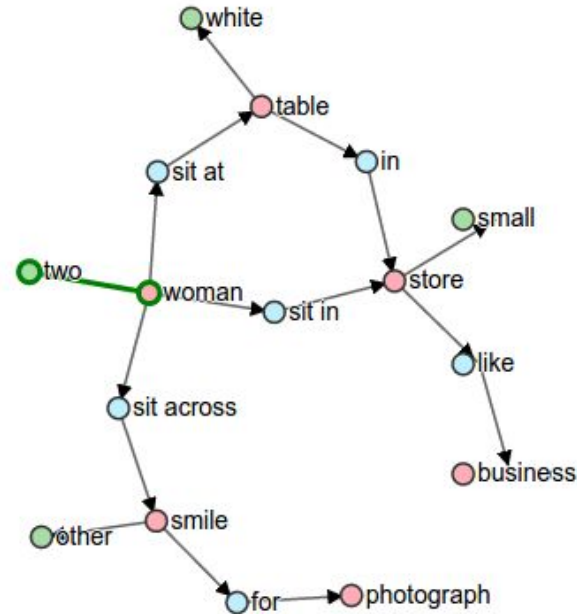# Relations (scene graphs and V&L)

"two women are sitting at a white table"

"two women sit at a table in a small store"

"two women sit across each other at a table smile for the photograph"
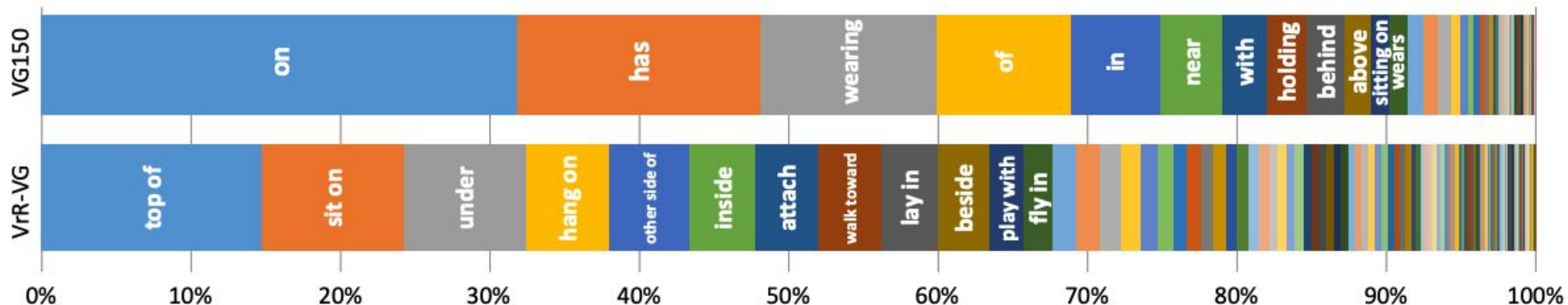
"two women sitting in a small store like business"

"two woman are sitting at a table"

Current datasets for relations

- Main benchmarks are Visual Genome (50 relations) and Open Images (1466)
- Not balanced
- Contains errors
- Misleading naming (many synonymous)

https://arxiv.org/pdf/1902.00313.pdf and https://arxiv.org/pdf/2104.01111.pdf

SOLUTION - CLEVR-Based relations dataset

- We need diagnostic tests to analyze our progress and discover shortcomings

- Existing benchmarks have strong biases that models can exploit to correctly answer questions without reasoning

- Conflate multiple sources of error, making it hard to pinpoint model weaknesses

- CLEVR is a diagnostic dataset that tests with minimal biases

- Unified distribution

- Out of distribution (artificial)
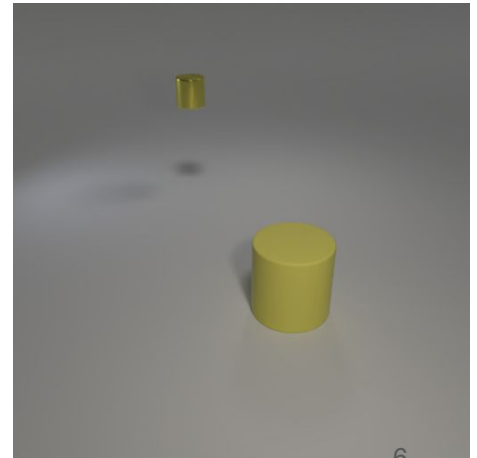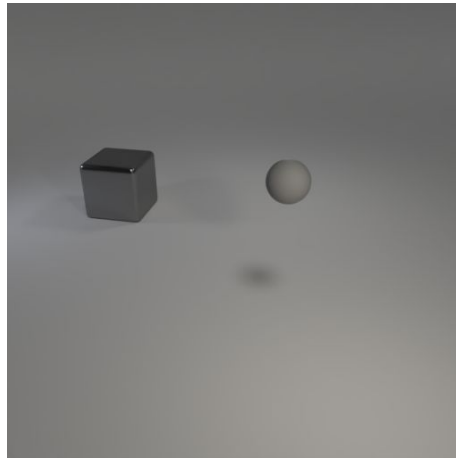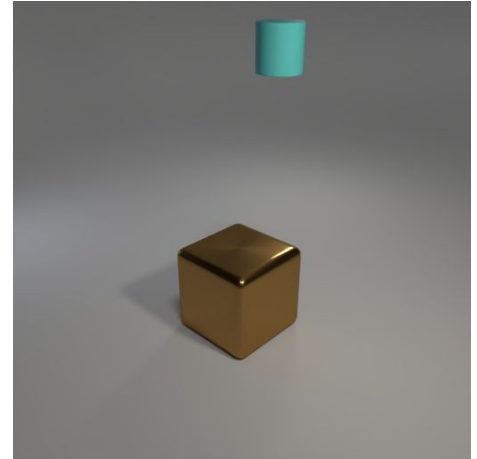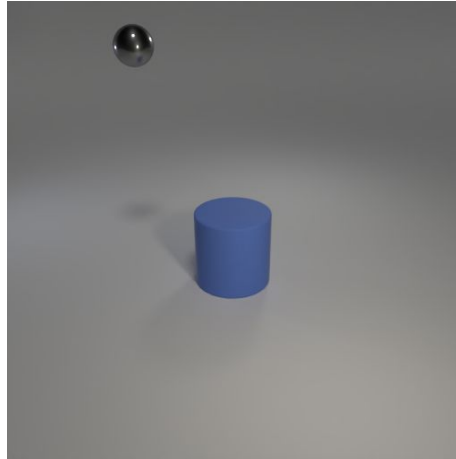
https://cs.stanford.edu/people/jcjohns/clevr/

Relations - Spatial
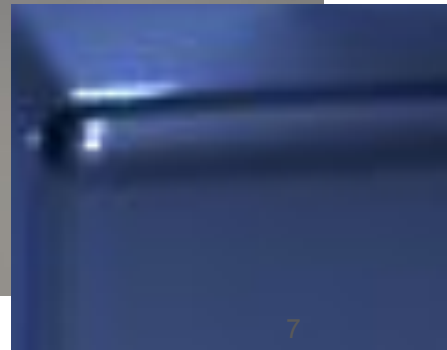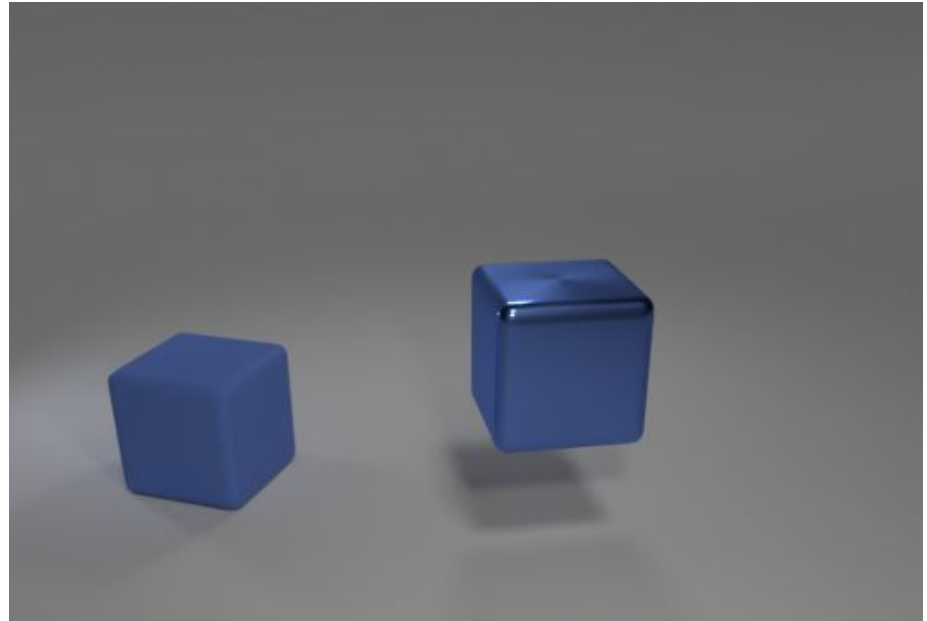
Bigger than (3 options)
Higher than (2 options)
Closer than (2 options)
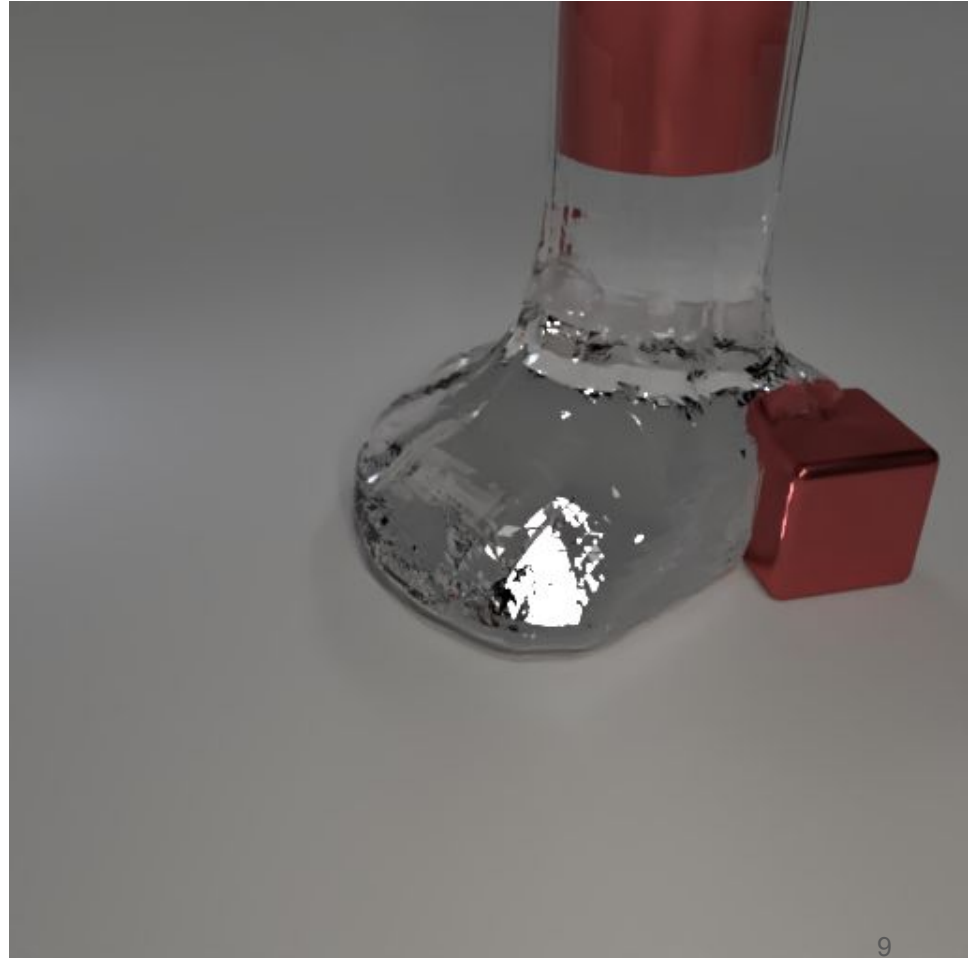Relative location - based on angle
(4 options)

Relations - Material

Shinier (two options)
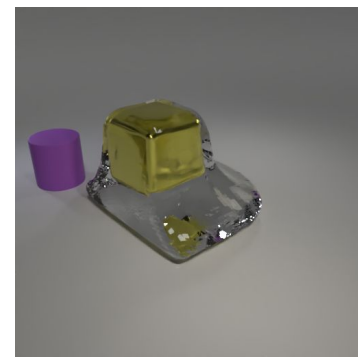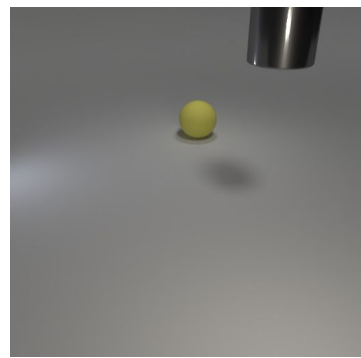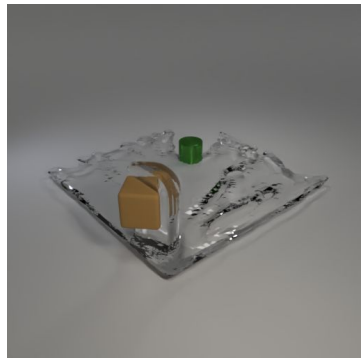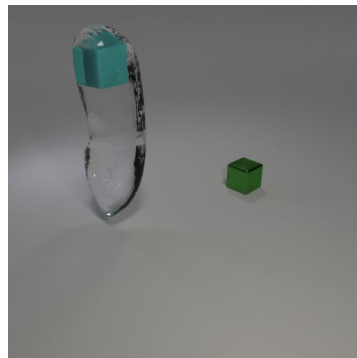
Relations - Liquid
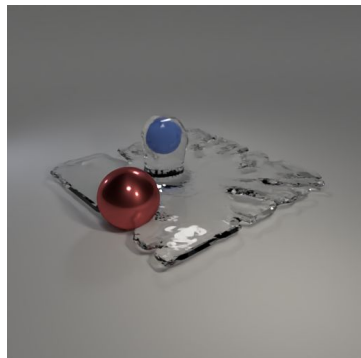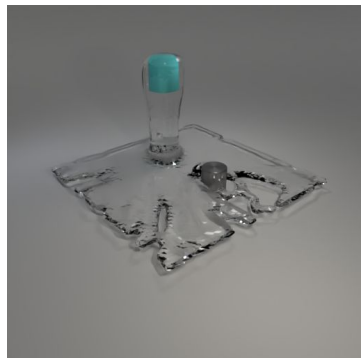
Flow from (3 options)

Relations - Liquid

Goo from (3 options)

# Relations - Examples   - 18 Relations in total.

Blender Implementation - Liquid Simulation

- Many Infinitesimal particles
- Conservation laws (mass, moment, **volume**, energy)
- Internal Forces (viscosity)
- External Forces (gravity, boundaries)
- Simplified Navier-Stokes solver
- FLIP
  - CPP implementation
  - Ships with blender since 2.9
  - Full API available
  - Can simulate liquid and gas
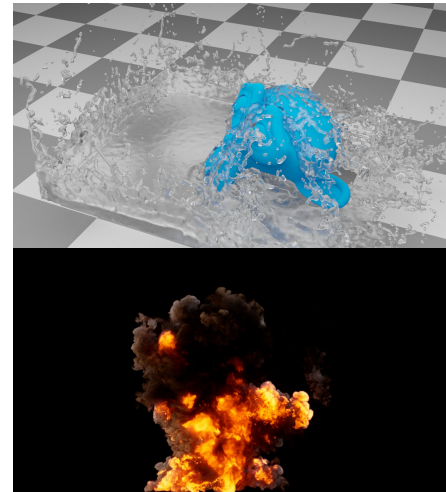
Blender Implementation - Scene Breakdown

1) Define liquid domain
2) Define sources, sinks and liquid properties
3) Run physical liquid simulation
4) Define scene objects and materials
5) Render
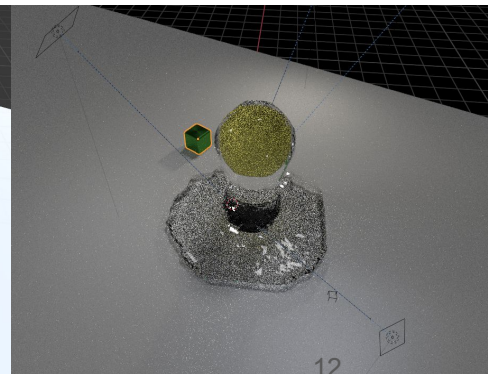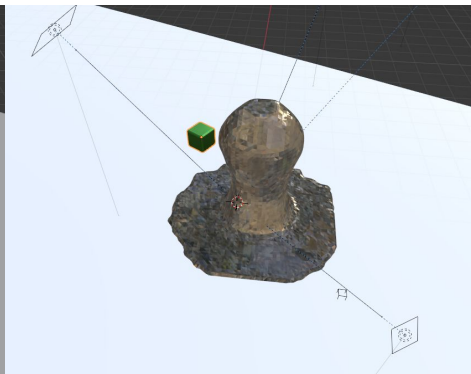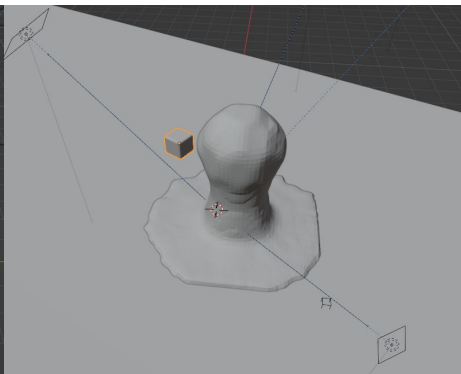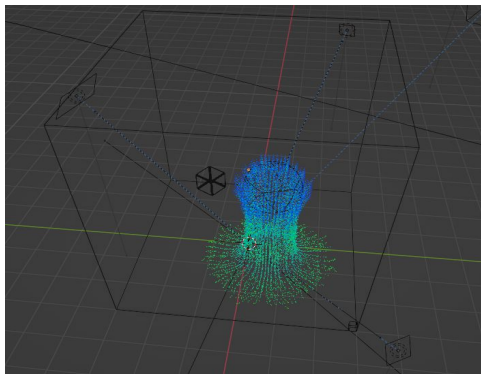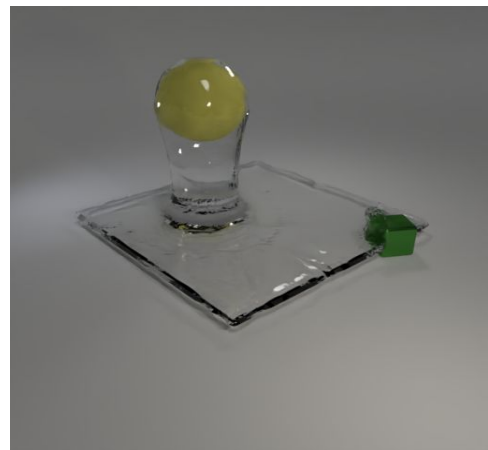
Challenges - Obscured Objects

We need to avoid obscured objects

- Render them "shadeless" - surprisingly difficult with physical engine
- Iterate over image pixels and assure # unique colors = 2 (+1)

Challenges - Extensions

- We want to allow extensions
  - New relations
  - New shapes
  - New materials
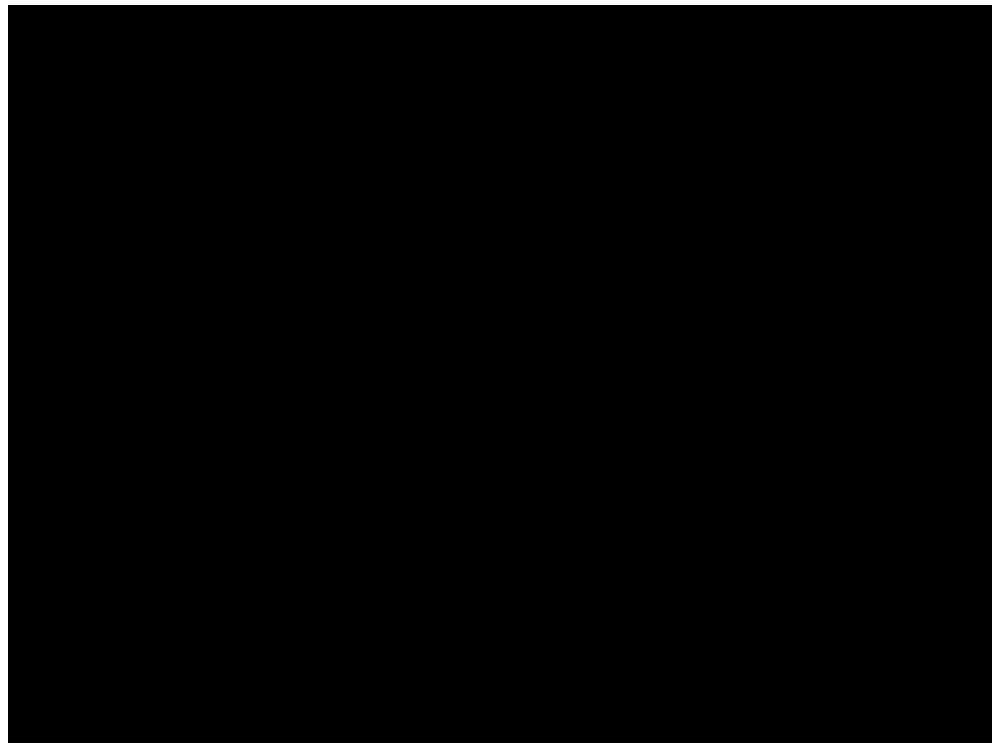 - New liquid properties
  - Support new blender versions

- Labels are encoded in a json format
- Materials are defined independently
- Migrated CLEVR to Blender 3.0 (and kept backward compatibility)

```
{
  "shapes": {
    "cube": "SmoothCube_v2",
    "sphere": "Sphere",
    "cylinder": "SmoothCylinder"
  },
  "colors": {
    "gray": [87, 87, 87],
    "red": [173, 35, 35],
    "blue": [42, 75, 215],
    "green": [29, 105, 20],
    "brown": [129, 74, 25],
    "purple": [129, 38, 192],
    "cyan": [41, 208, 208],
    "yellow": [255, 238, 51]
  },
  "materials": {
    "rubber": "Rubber",
    "metal": "MyMetal"
  },
  "liquid_materials": {
    "water": "Water",
    "lava": "Lava",
    "smoke": "Smoke"
  },
  "sizes": {
    "large": 0.7,
    "small": 0.35
  }
}
```

Challenges - Hyper Parameters

Vast amounts of hyper parameters
- Liquid Simulation
- Rendering & Lighting
- Randomization

We manually selected what seemed "ok".

Future directions might be to run a grid parameter search
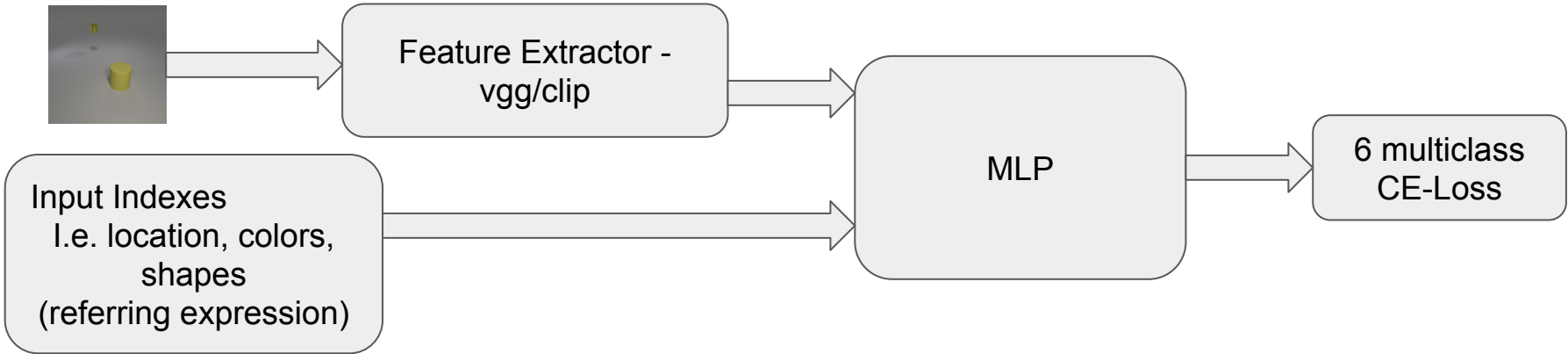
Challenges - Render Times

Rendering 5K images & their labels was infeasible on home PC

- Rendering time
  - We used crowd-render and the Bermano01 server (8 x GeForce RTX 2080 Ti)

- Physical simulation
  - Decreased time of simulation and increased liquid particle size
  - Compromised on liquid quality
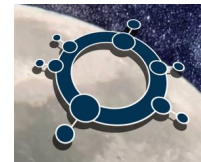

crowd-render.com

Baseline Model

Benchmark Results - Accuracy

Relation type \model	random	vgg-features	Clip ViT	Clip RN50	vgg-no-location-input
Greater	0.333	0.87	0.47	0.48	0.858
Higher	0.5	1.0	1.0	1.0	0.996
Sparklier	0.5	0.874	0.52	0.49	0.89
Relative Location (by angel)	0.25	0.975	0.98	0.98	0.84
Liquid	0.2	0.993	0.96	0.95	0.994
Closer than	0.5	0.88	0.82	0.80	0.84
#Average	0.38	0.932	0.784	0.783	0.903

Examples of accuracy graphs
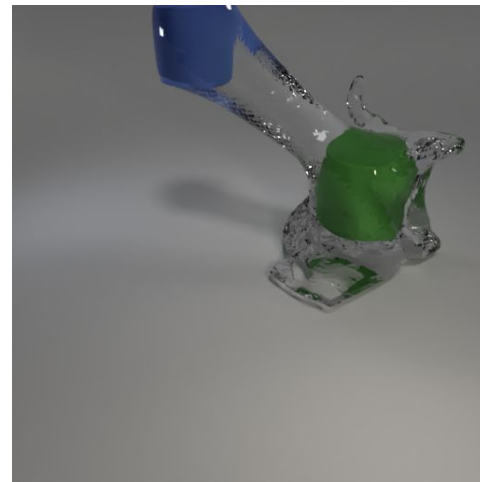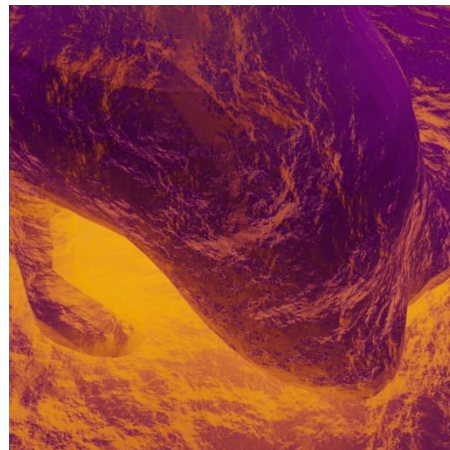

val_acc
tag: val_acc
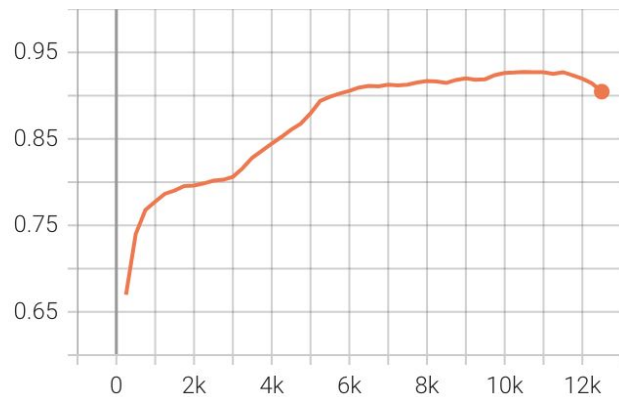
g: 4_val_acc
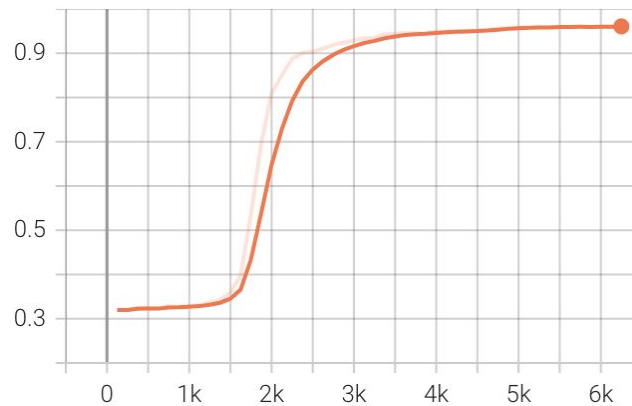
Discussion & Future Work

- CLEVER-ER allows better understanding of how networks deal with relationships

- Incorporate less trivial liquid relationships
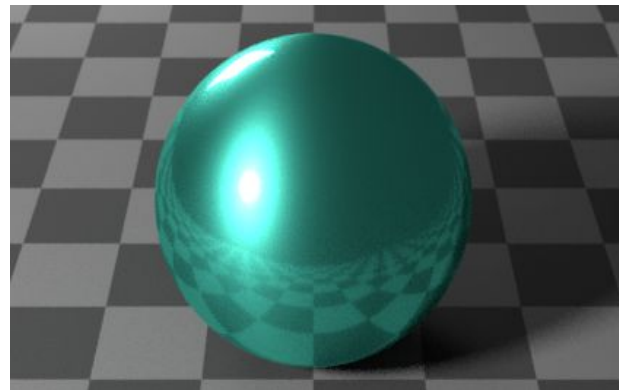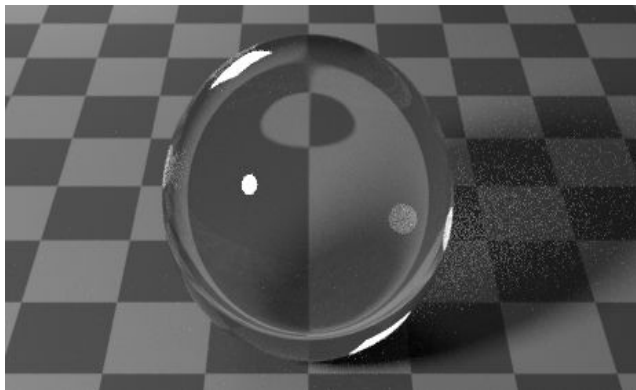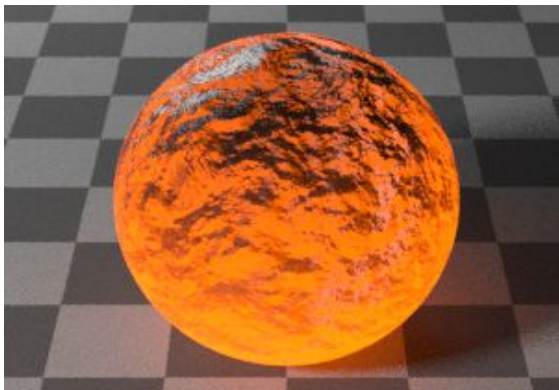 - Sprinkle
  - Blow
  - Drip
  - Splash

Discussion & Future Work

Render of liquid lava
from our code

- Extend benchmark to V&L tasks as VQA
- Extend material list
  - Lava
  - Smoke
  - More

Feel free to check it out ! (and contribute… or to star us)

https://github.com/yoterel/CLEVR-ER