ShapeStacks

Learning Vision-Based Physical Intuition for Generalised Object Stacking

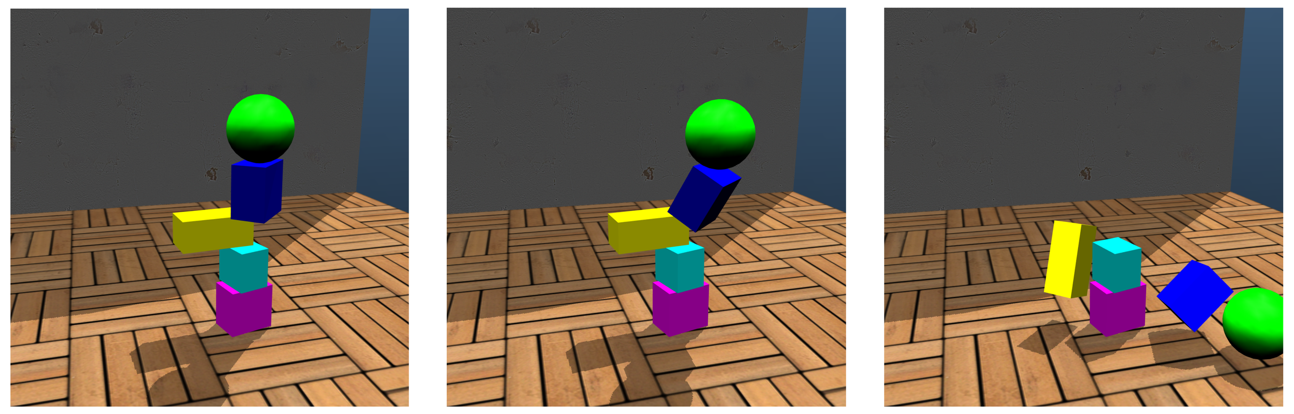
Oliver Groth, Fabian Fuchs, Ingmar Posner, Andrea Vedaldi

Applied Artificial Intelligence Lab, Oxford Robotics Institute

Visual Geometry Group, Department of Engineering Science, University of Oxford

University of Oxford

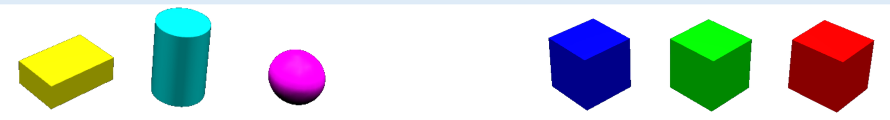
Introduction



ShapeStacks is a **simulation dataset** for the learning of **vision-based physical intuition**.
We present a model for **visual stability prediction** and demonstrate its **applicability in an object stacking task**.

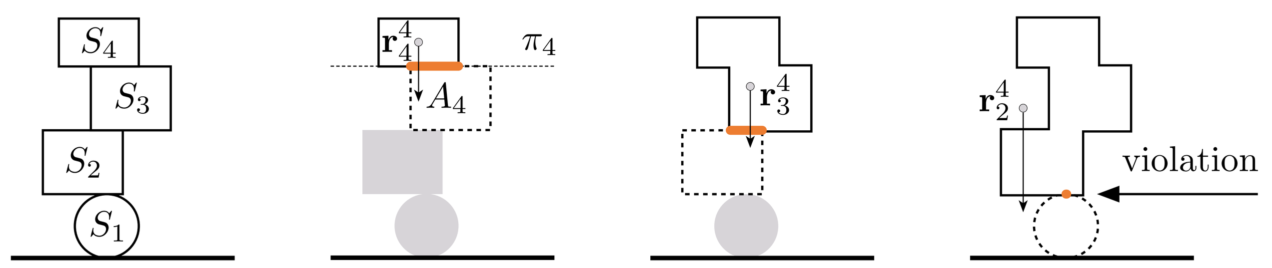
Dataset

The ShapeStacks dataset is openly available at **http://shapestacks.robots.ox.ac.uk/**



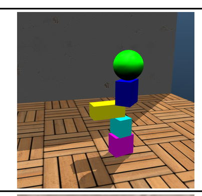
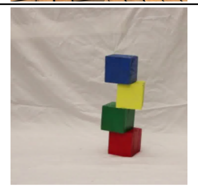
Height	CCS (# Scenarios) Train	CCS (# Scenarios) Test	Cubes (# Scenarios) Train	Cubes (# Scenarios) Test
h = 2	1,626	286	2,040	360
h = 3	2,992	528	2,040	360
h = 4	2,084	368	1,890	332
h = 5	822	144	1,546	272
h = 6	234	40	1,250	220
# Scenarios	7,758	1,366	8,766	1,544

Each scenario features a unique object stack and is available as **MuJoCo [1] simulation environment and RGB-D image sequence.**



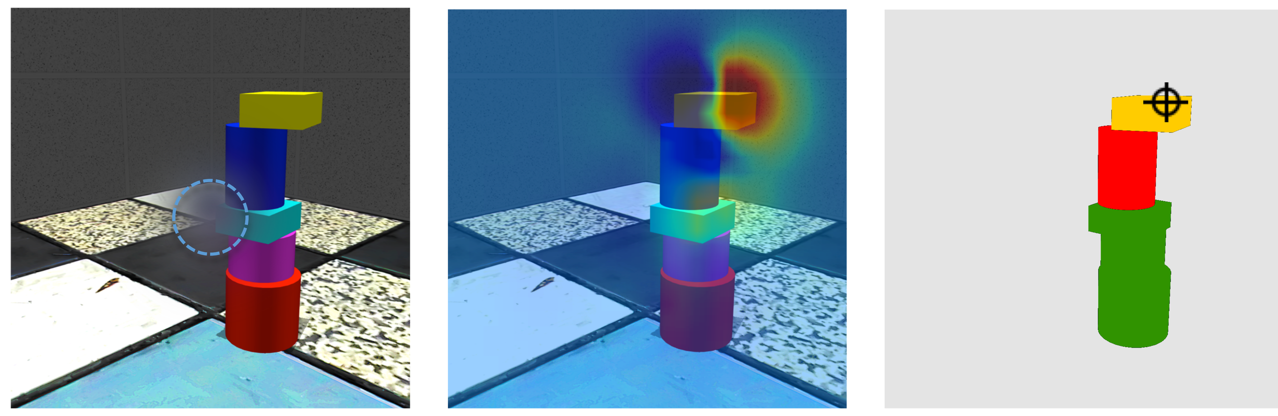
Segmentation maps annotate **each object's mechanical contribution to tower stability.**

Stability Prediction

		ShapeStack Inception [2]	PhysNet [3] (trained on RGB images and segmentation maps)	VDA [4] (requires physics simulator during prediction)
SIM		**84.9%**	N/A	N/A
REAL		74.7%	66.7%	**75.0%**

Our model achieves **state-of-the-art stability prediction accuracy** on simulated and real images [2] with a simplified training and inference setup compared to related work.

Structure Analysis



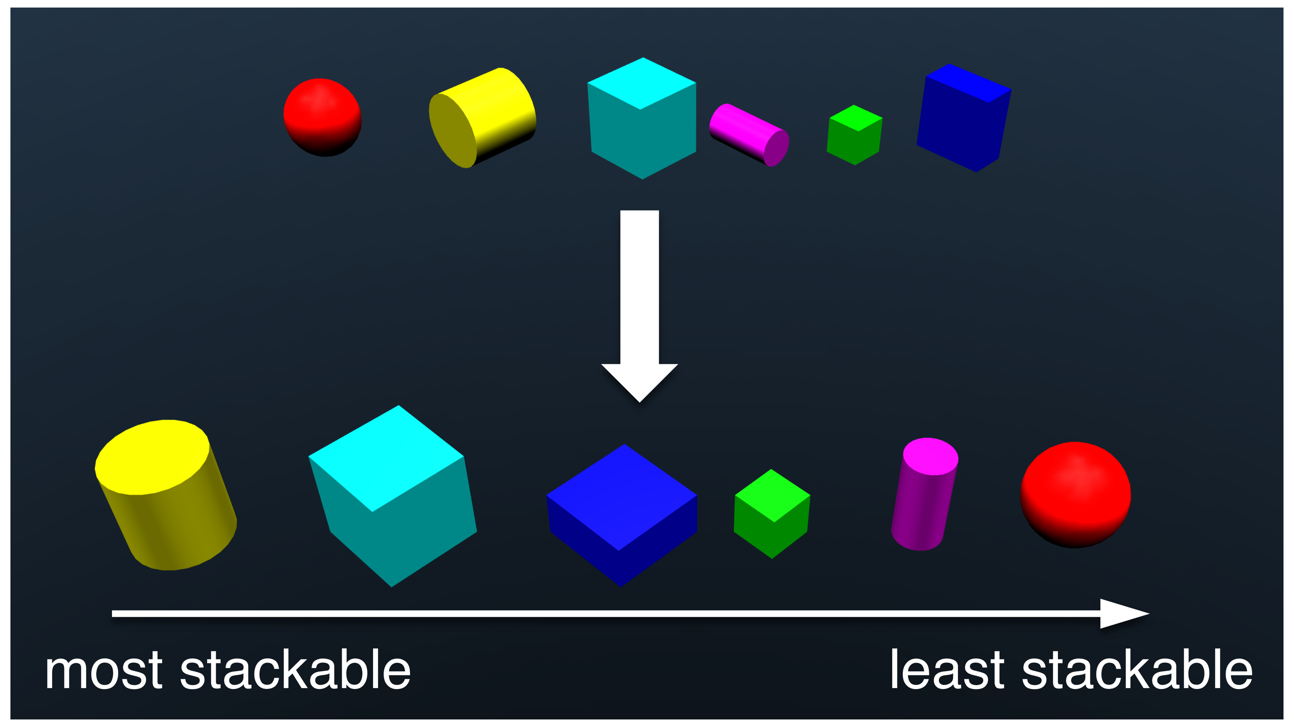
An **attention analysis** of the stability predictor was conducted via a blur study and compared to the annotation of the **stack mechanics** via the ground truth segmentation maps.

	Violating object	First object to fall	Rest of tower	Back-ground
VCOM & VPSF	38.9%	29.3%	11.4%	20.4%
VCOM	32.7%	30.8%	13.8%	22.7%
VPSF	52.1%	26.3%	6.3%	15.4%
Random chance	1.6%	1.9%	3.5%	93.0%
Random in tower	19.3%	22.9%	42.2%	14.5%

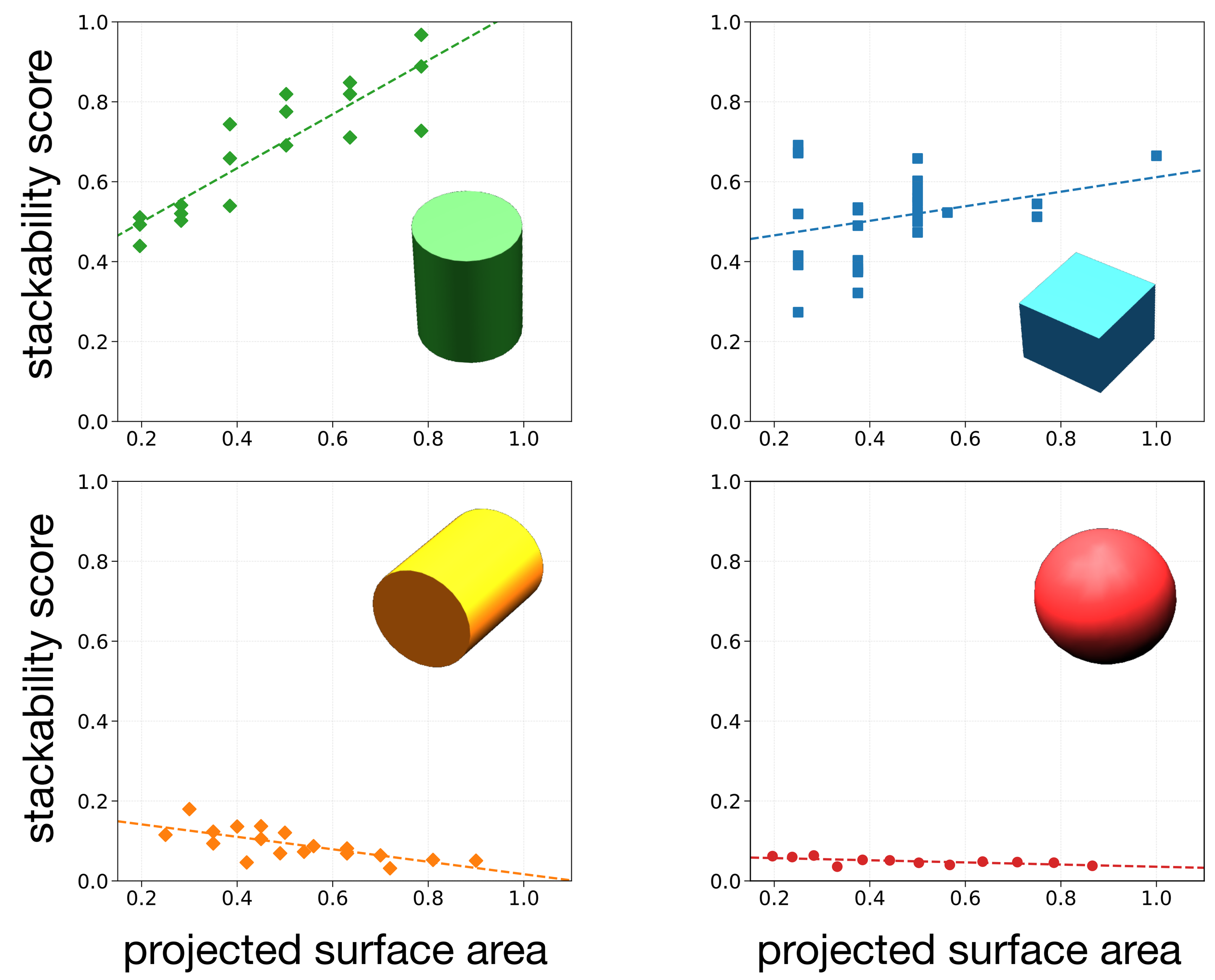
VCOM = tower with COM violation; VPSF = tower with non-planar support interfaces

Significant **focus on mechanically relevant regions** implies an intuitive understanding of the ***centre-of-mass principle (COM)***.

Stackability Score

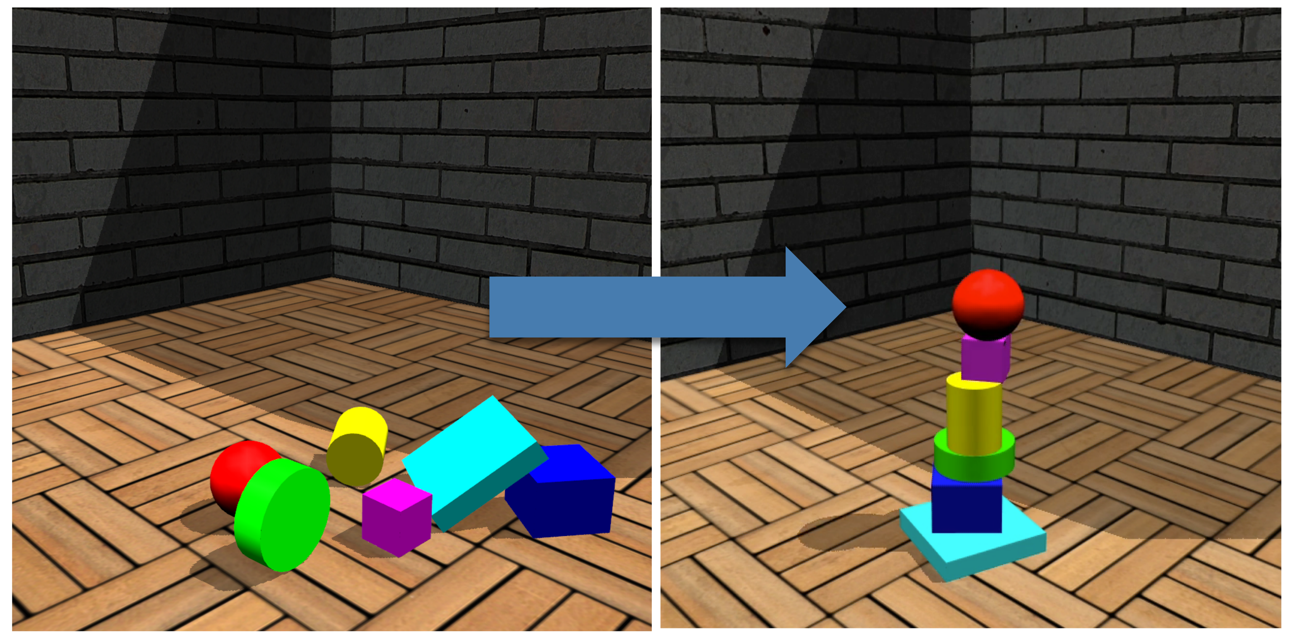


Ranking of random object set with respect to ***stackability***, i.e. the stacking order of objects which affords the highest possible tower.

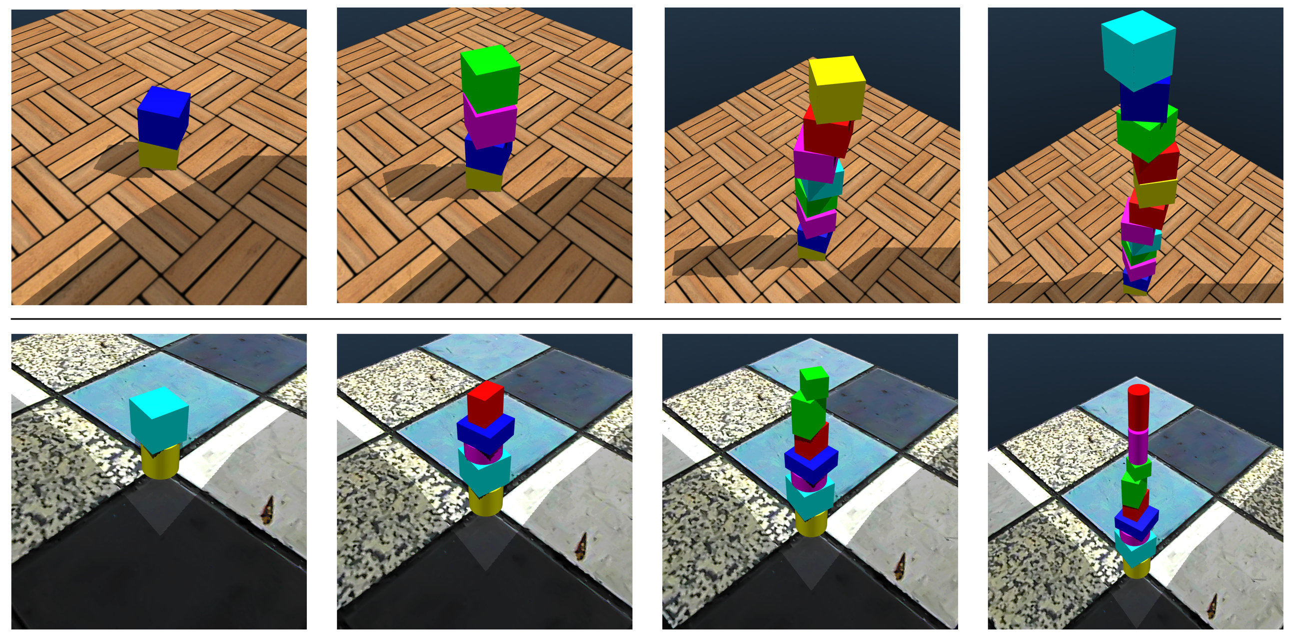


Stackability indicates the **capability of an object to support** the others in the set.
The score **correlates positively with the projected surface area** as long as the support interface is planar.

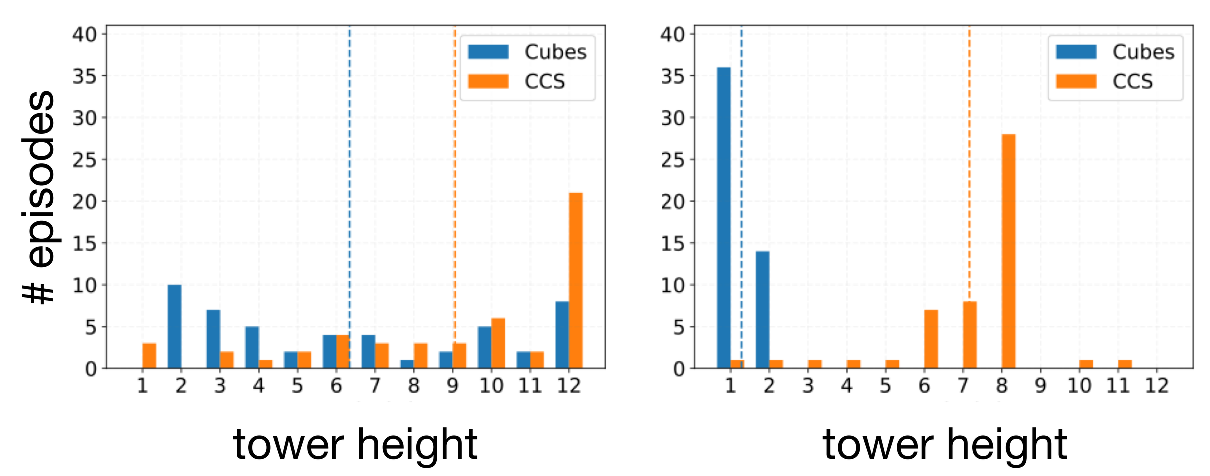
Stacking Task



Ranking, orientation and placement of objects in a tower based on the stability prediction via a **simulated annealing process**.



The **stability predictor** is used to **determine stable stacking positions.** Predictors trained on diverse towers (**C**uboids-**C**ylinders-**S**pheres) perform best building stacks with an **average height of eight objects**.



Dataset, code and models available!
RGB images, segmentation & depth maps
MJCF environments & simulation code
ShapeStacks scenario generator
pre-trained models for stability predictor

References
[1] Todorov, E., Erez, T., Tassa, Y.: *MuJoCo: A physics engine for model-based control.* IROS, 2012.
[2] Szegedy, C., Ioffe, S., Vanhoucke, V., Alemi, A. A.: *Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning.* AAAI, 2017.
[3] Lerer, A., Gross, S., Fergus, R.: *Learning Physical Intuition of Block Towers by Example.* ICML, 2016.
[4] Wu, J., Lu, E., Kohli, P., Freeman, W. T., Tenenbaum, J. B.: *Learning to See Physics via Visual De-animation.* NIPS, 2017.

research funded by ERC 677195-IDIU
European Research Council
Established by the European Commission