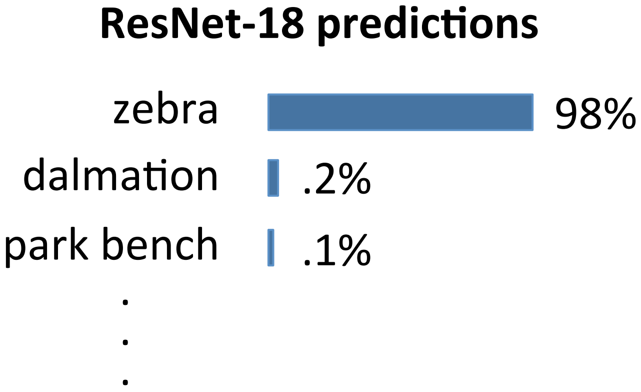
Causal and compositional generative models in online perception

Michael Janner
with Ilker Yildirim, Mario Belledonne, Christian Wallraven, Winrich Freiwald, Joshua Tenenbaum
MIT

July 28, 2017

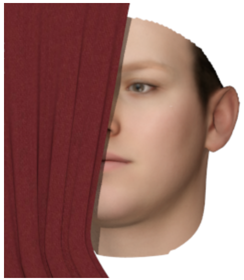
London, UK



ResNet: He et al. *CVPR* 2016.

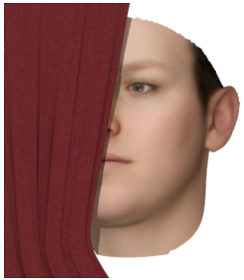
Outline

1. Occluded face perception with causal and compositional models

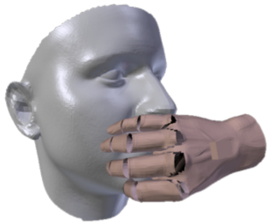


Outline

1. Occluded face perception with causal and compositional models

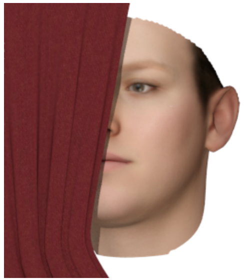


2. Automatic training-free vision-to-touch crossmodal transfer

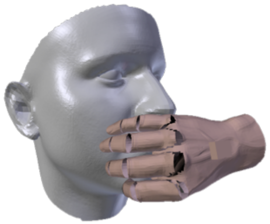


Outline

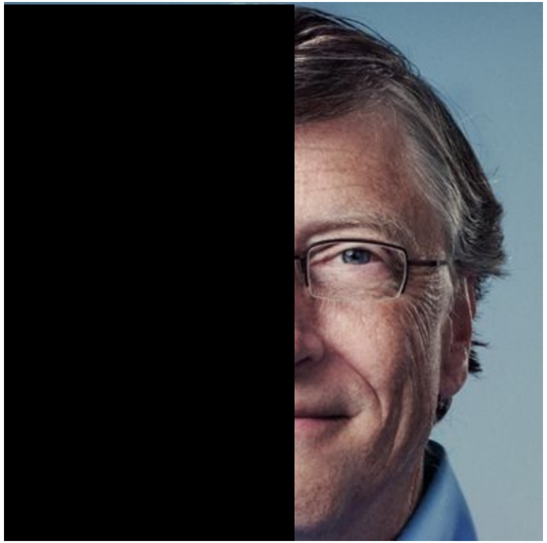
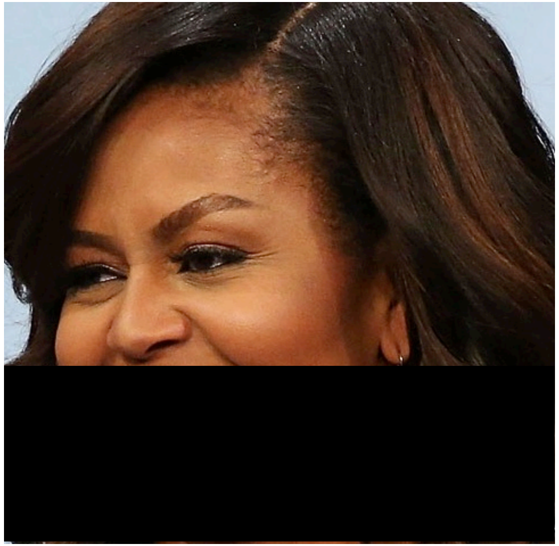
1. Occluded face perception with causal and compositional models



2. Automatic training-free vision-to-touch crossmodal transfer



3. Learning visual causal models for generic object categories


The generative model

Intrinsic

Shape S;
Texture T

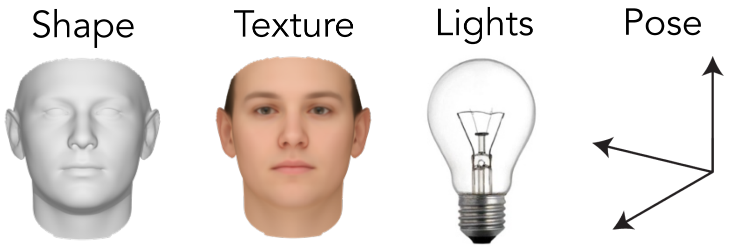
Shape

Texture

Face Morphable Model: Blanz, Vetter. *SIGGRAPH* 1999.

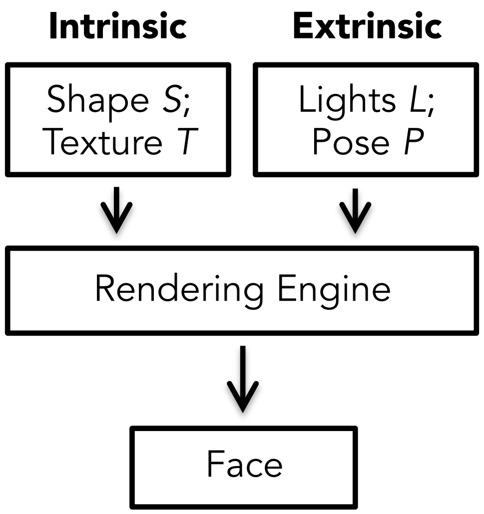
The generative model

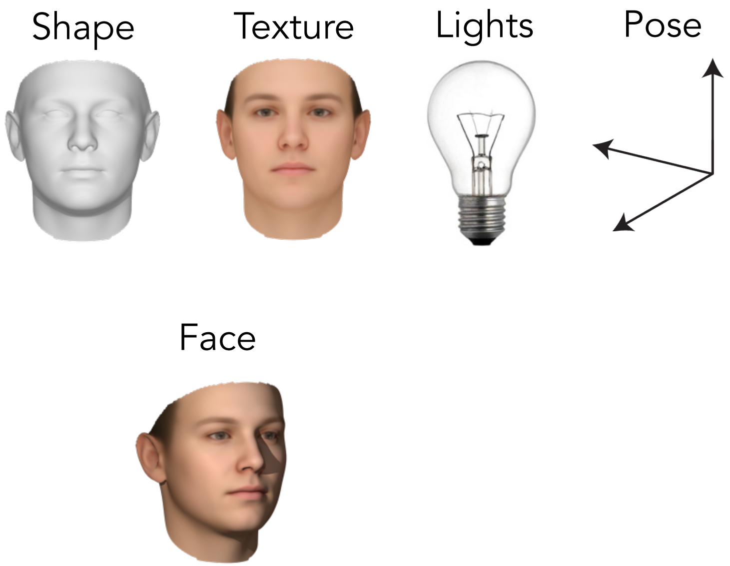
Intrinsic	**Extrinsic**
Shape S; Texture T	Lights L; Pose P



Face Morphable Model: Blanz, Vetter. *SIGGRAPH* 1999.

The generative model





Face Morphable Model: Blanz, Vetter. *SIGGRAPH* 1999.

The generative model

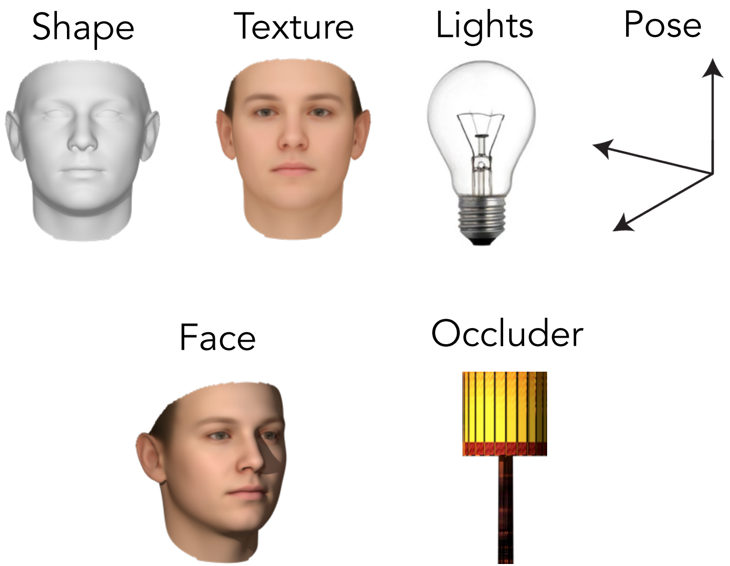
Intrinsic | **Extrinsic**

Shape S; Texture T → Rendering Engine

Lights L; Pose P → Rendering Engine

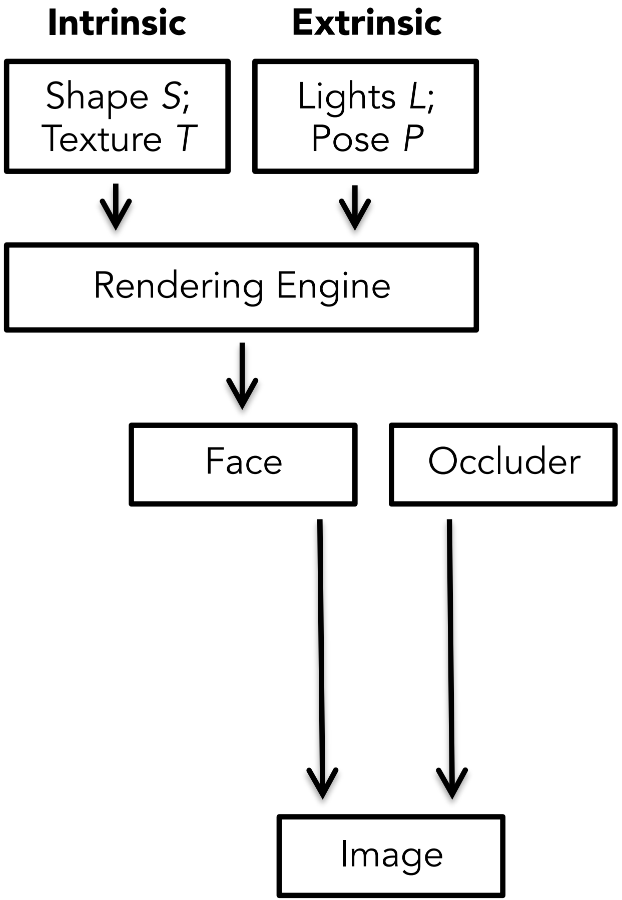
Rendering Engine → Face

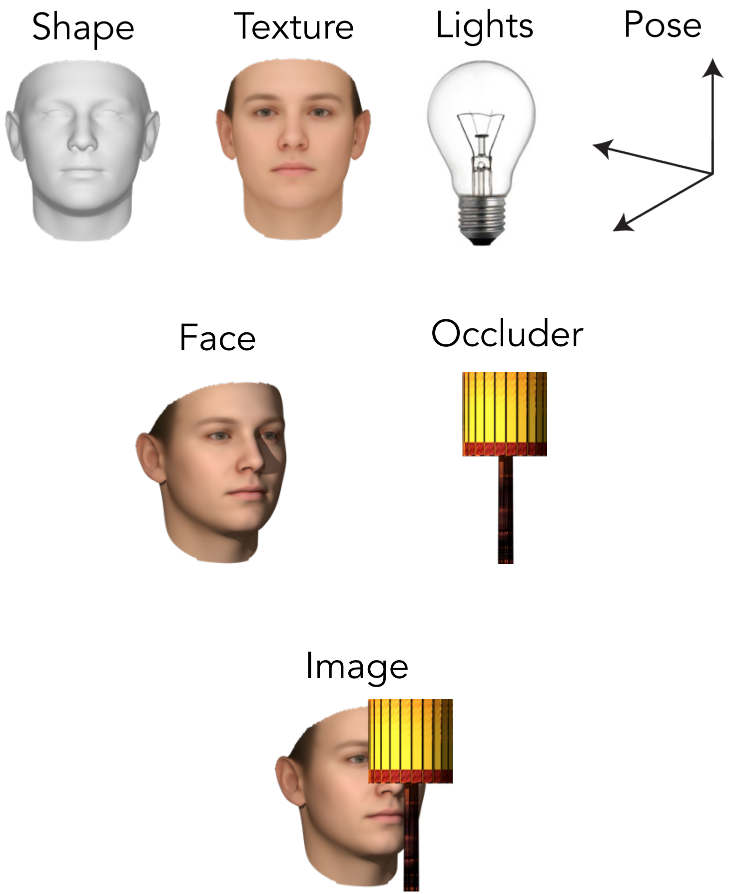
Face | Occluder



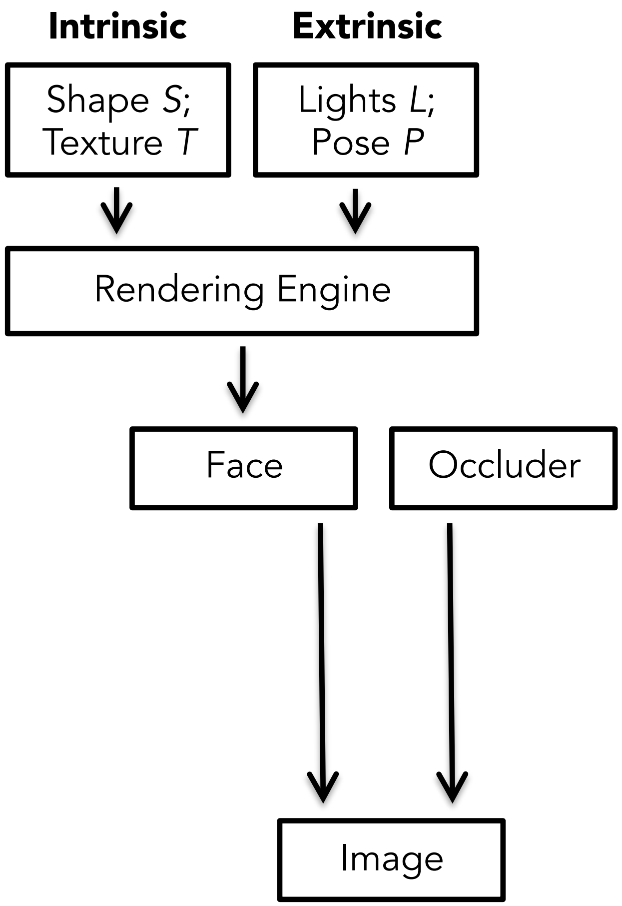
ShapeNet: Angel X. Chang, et al. arXiv:1512.03012

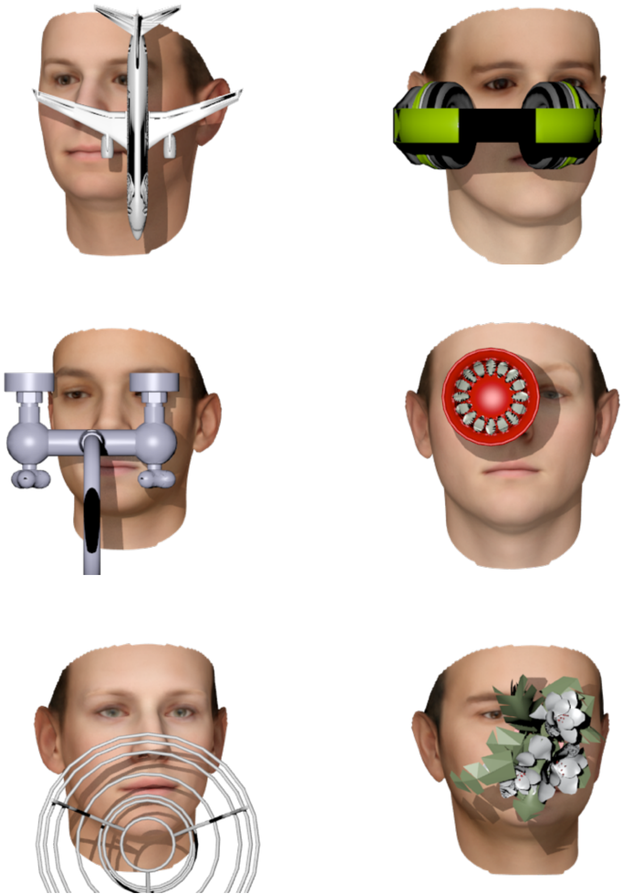
The generative model



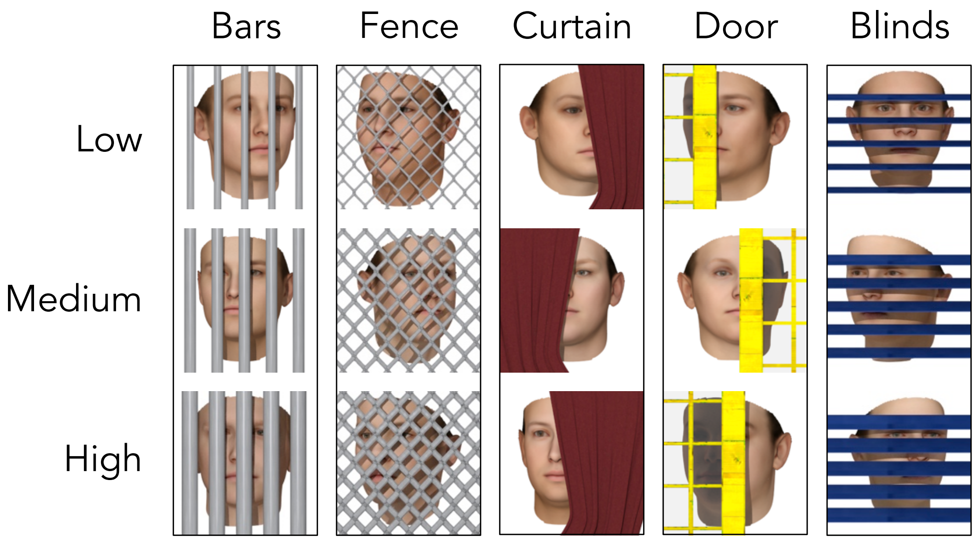


Samples from the generative model





Test occluders



- Occlude face in stripes, mesh patterns, and large patches
- Three levels of occlusion, ranging from 15-55% of face covered

Occluded face perception task

Occluded face perception task

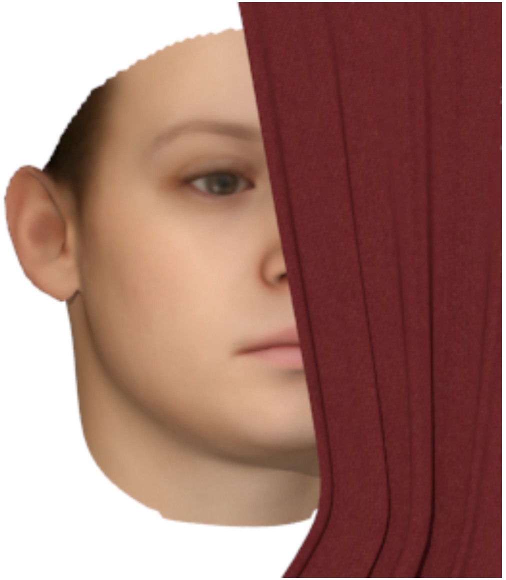
Occluded face perception task

Occluded → Unoccluded

Same

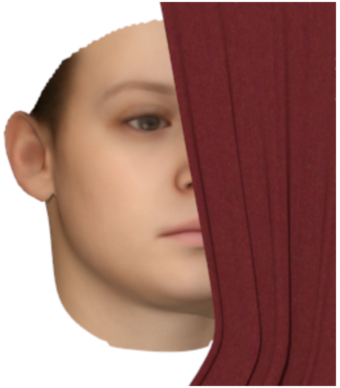
Occluded face perception task

Occluded face perception task



Occluded face perception task

Unoccluded → Occluded



Different

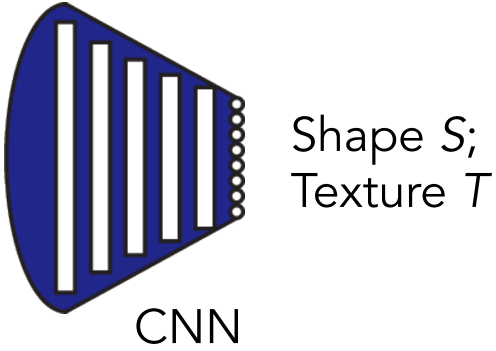
Behavioral results

Accuracy

	Occlusion Level		
	Low	Medium	High
Occluded → Unoccluded	.77	.73	.67
Unoccluded → Occluded	.78	.76	.70

Chance: .5

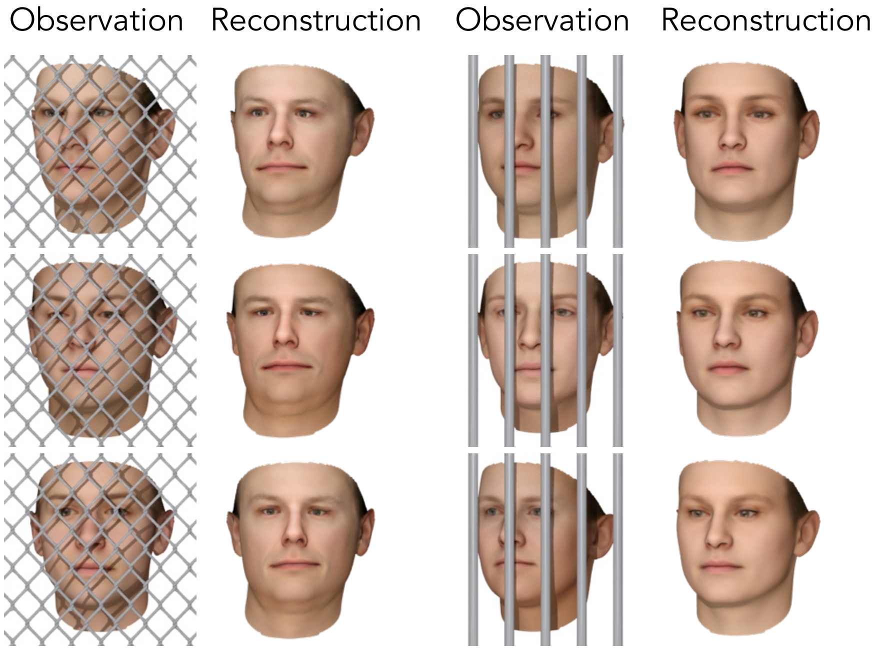
Naïve model: Learn to ignore the occluder



Predict intrinsic and extrinsic face parameters directly.

Can model become invariant to all types of occluders?

Naïve model: Learn to ignore the occluder

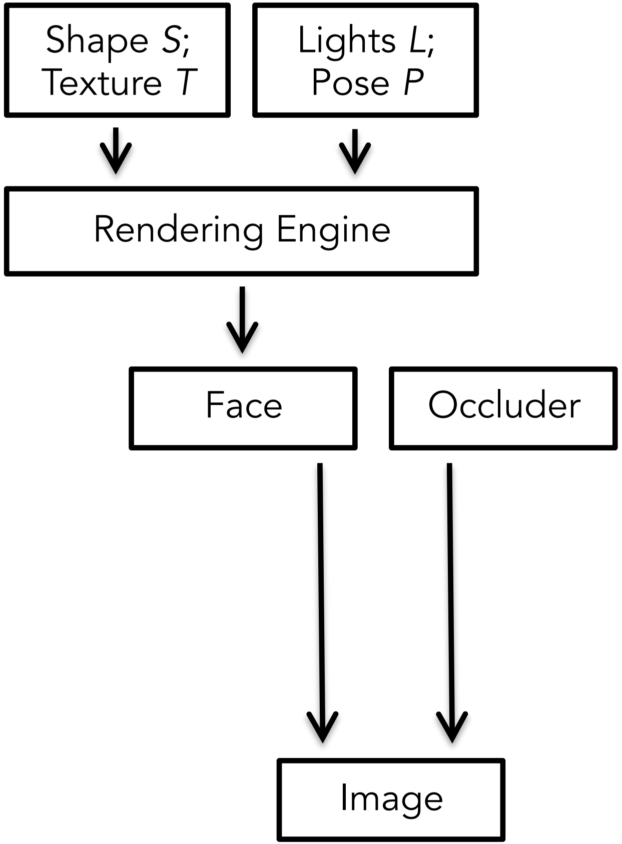


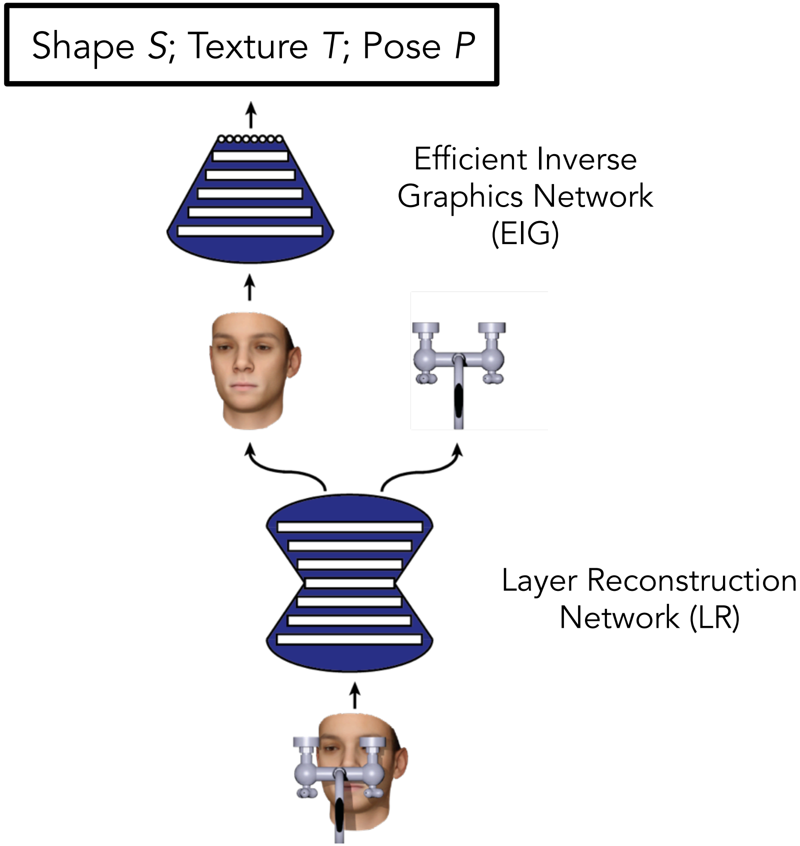
Predict intrinsic and extrinsic face parameters directly.

Can model become invariant to all types of occluders? **No.**

Test occluders heavily influence face predictions.

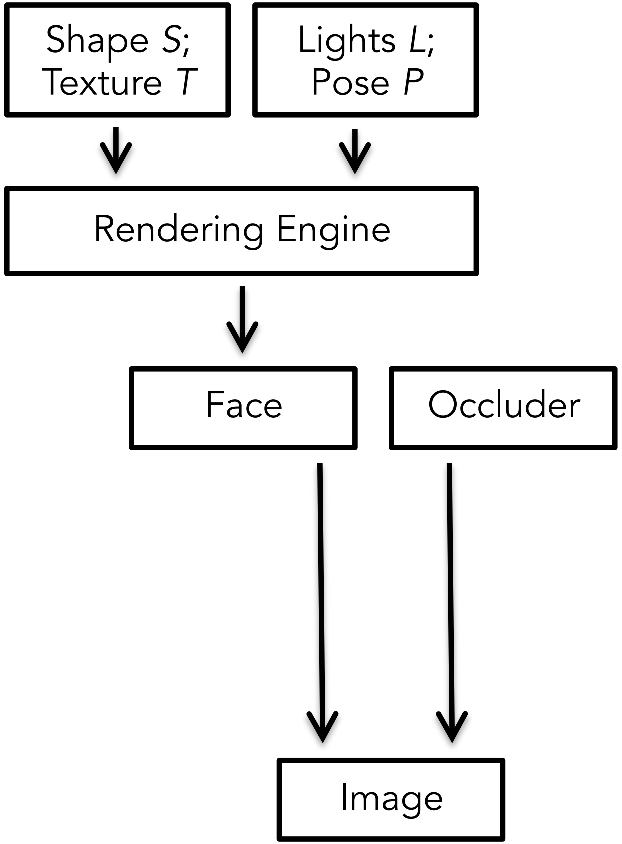
Inverse graphics: Inverting the generative model

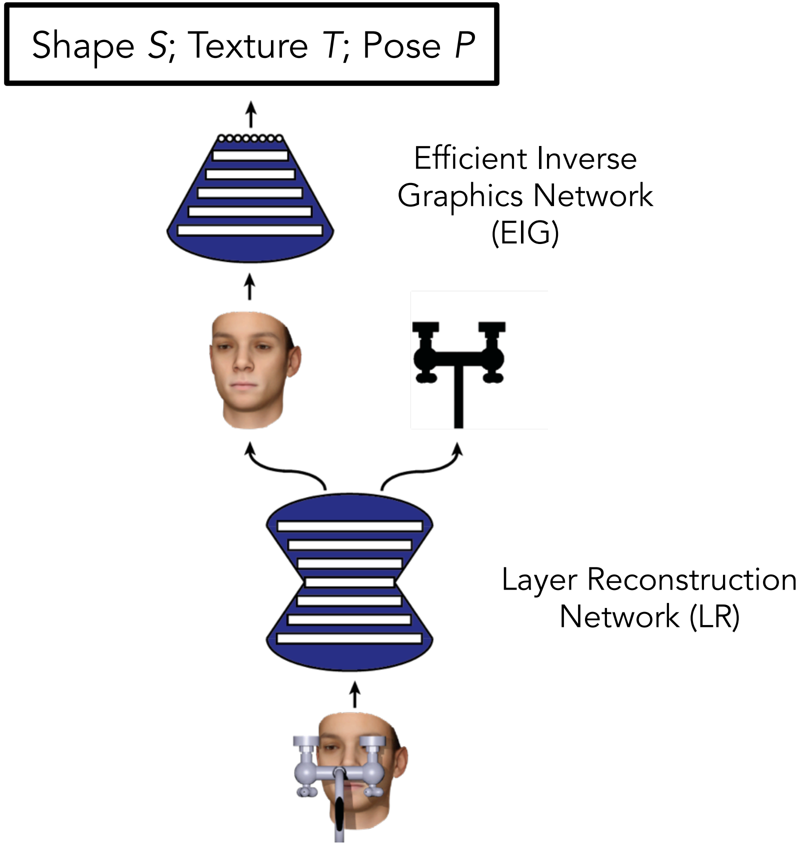




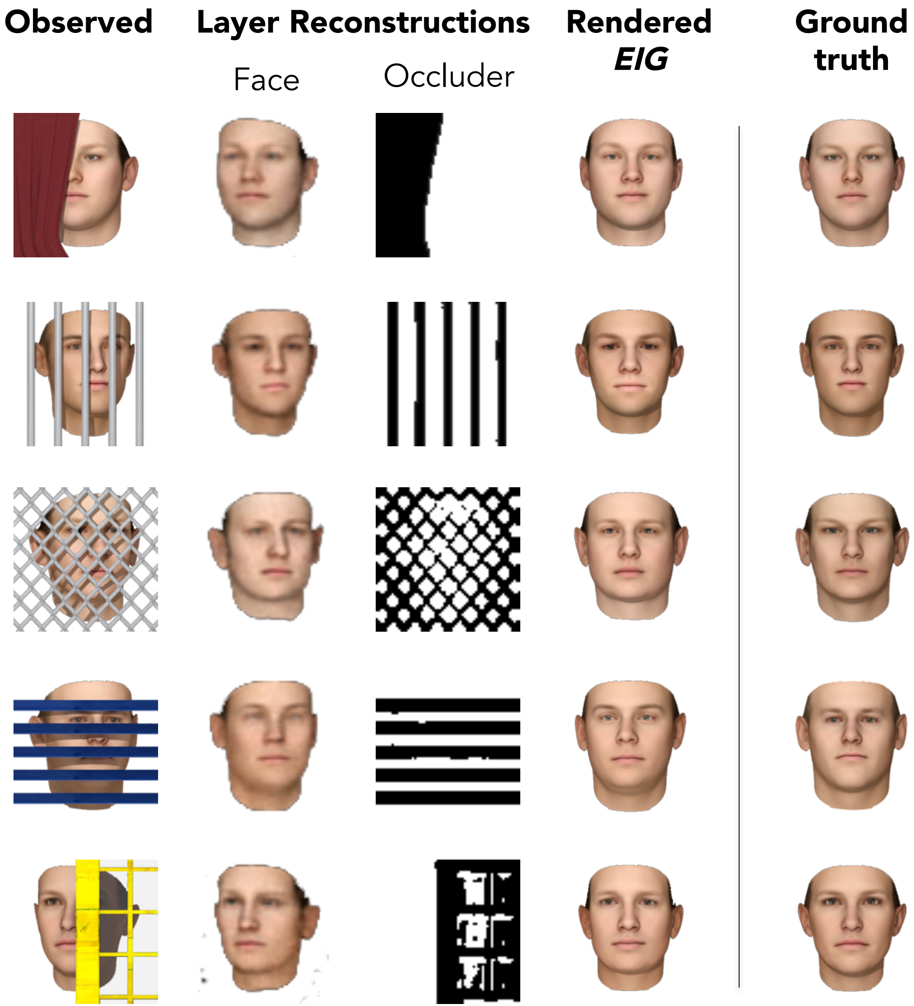
[Also see: Egger, *et al. Occlusion-aware 3D Morphable Face Models*. BMVC 2016.]

Inverse graphics: Inverting the generative model





[Also see: Egger, *et al. Occlusion-aware 3D Morphable Face Models*. BMVC 2016.]

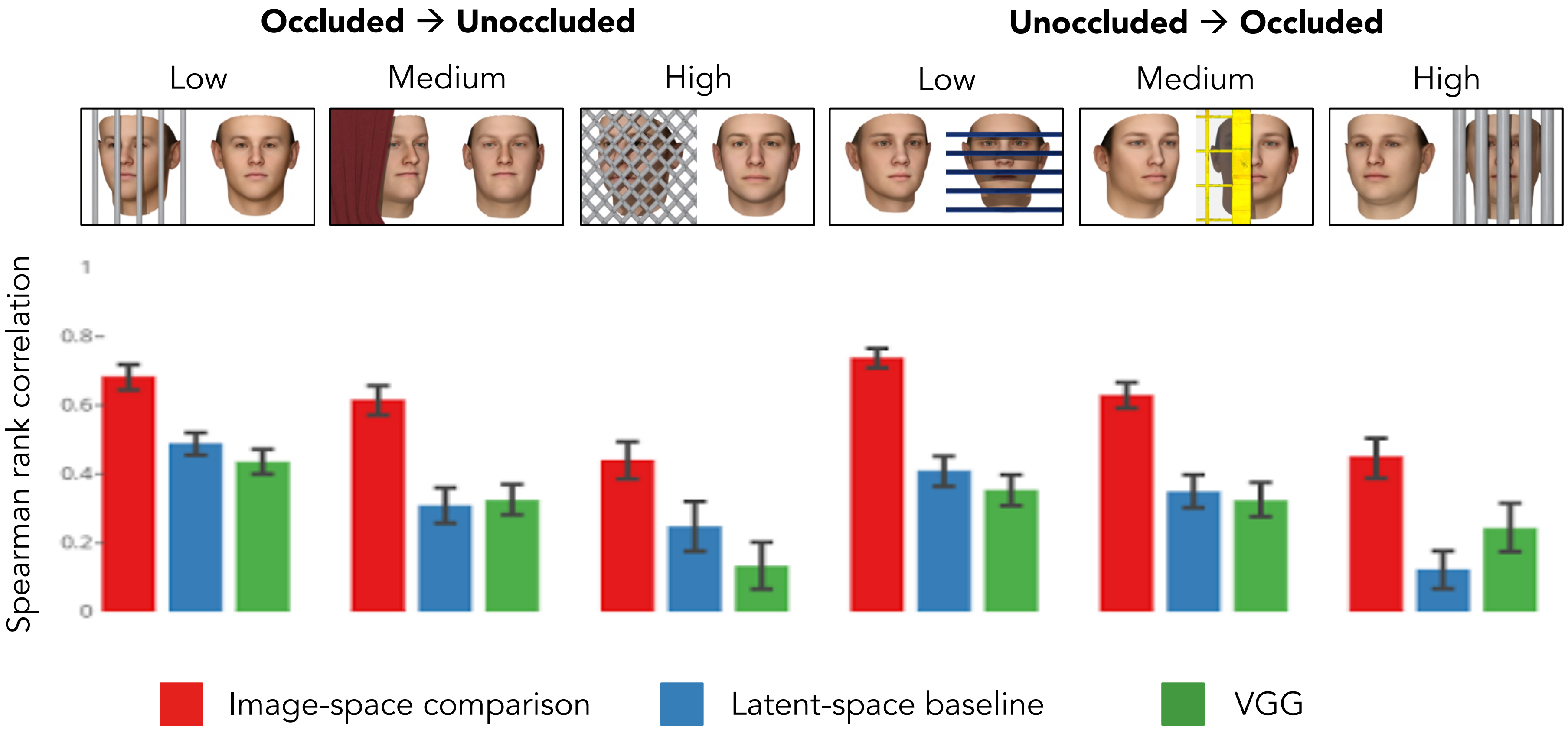


Modeling causes allows for:

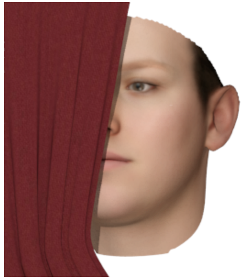
(1) face predictions that are invariant to occluders

(2) better generalization to unseen occluders

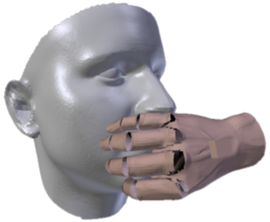
Prediction human judgments



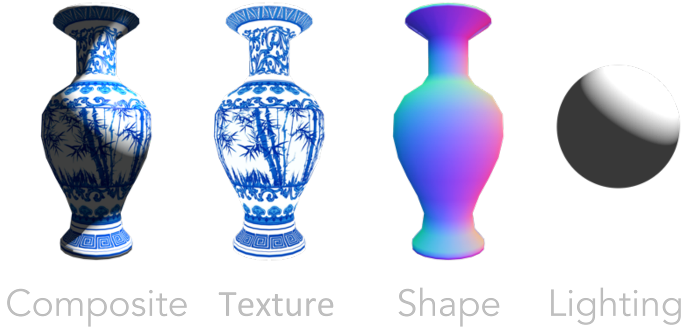
1. Occluded face perception with causal and compositional models



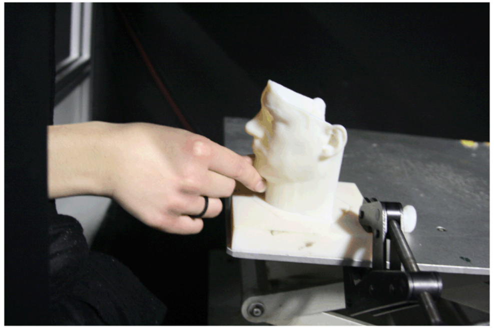
2. Automatic training-free vision-to-touch crossmodal transfer



3. Learning visual causal models for generic object categories



Vision-to-touch transfer



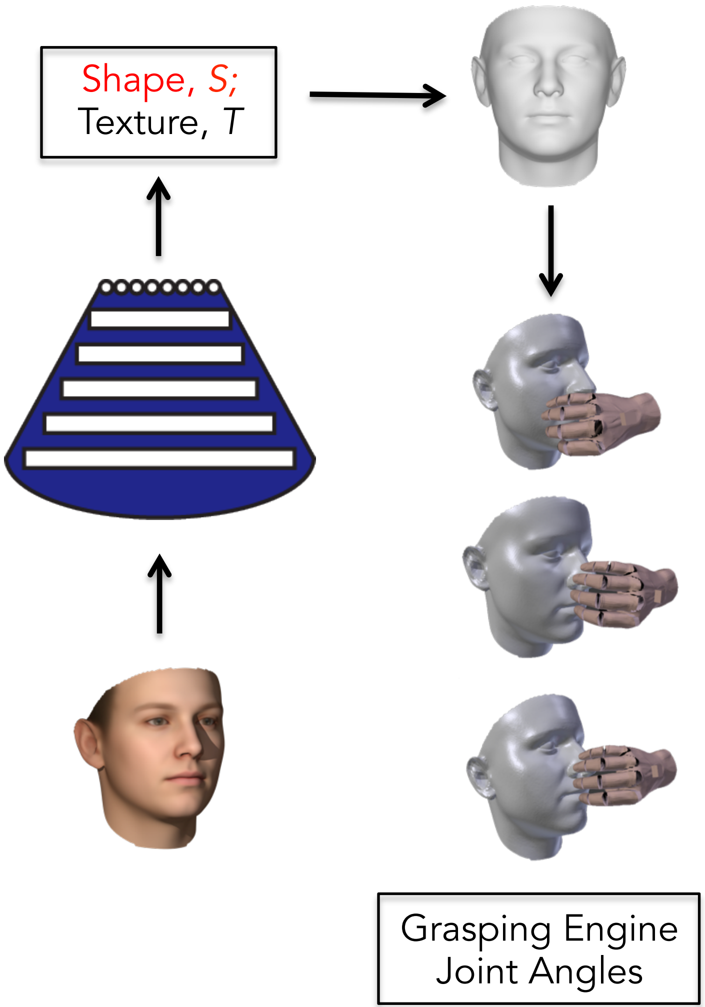
Study **Test**

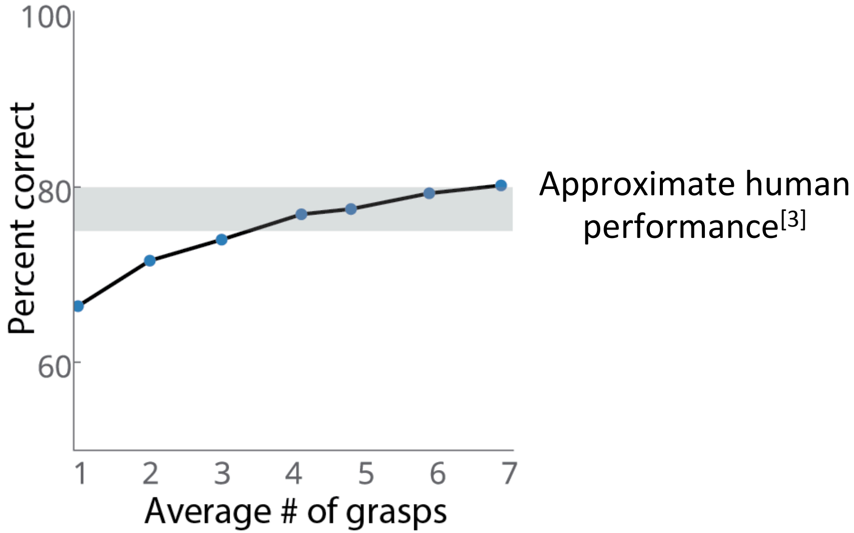
Shape is explicitly represented

Immediate transfer to haptic domain

Can make judgments based on grasping joint angles

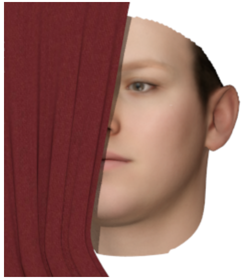
Vision-to-touch transfer



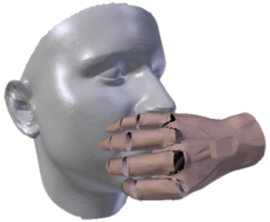


[3] Dopjans et al, *IEEE Transactions on Haptics*, 2000

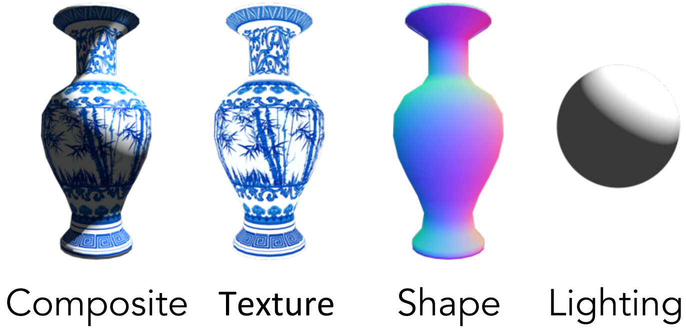
1. Occluded face perception with causal and compositional models



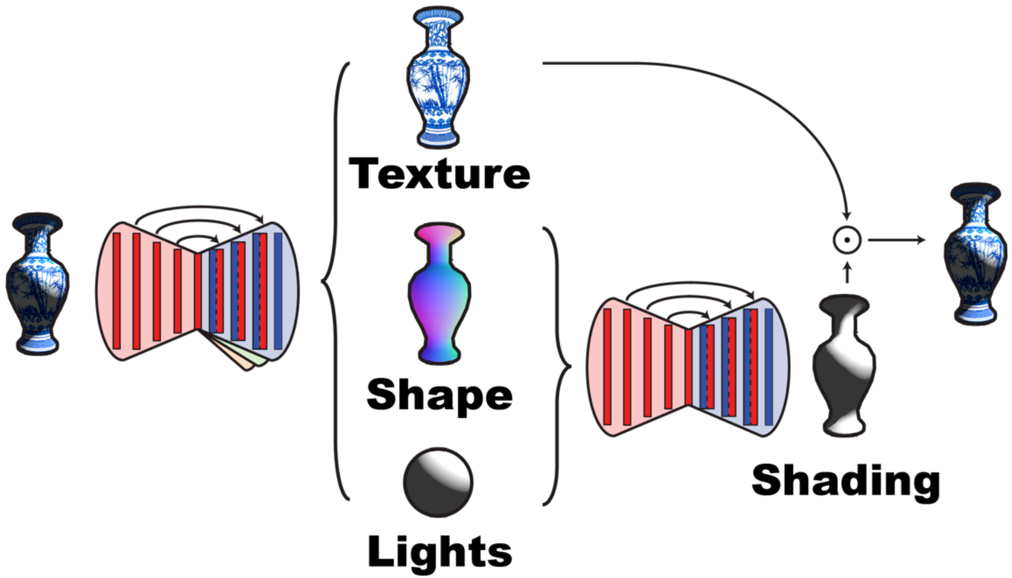
2. Automatic training-free vision-to-touch crossmodal transfer



3. Learning visual causal models for generic object categories



Causal models of generic objects

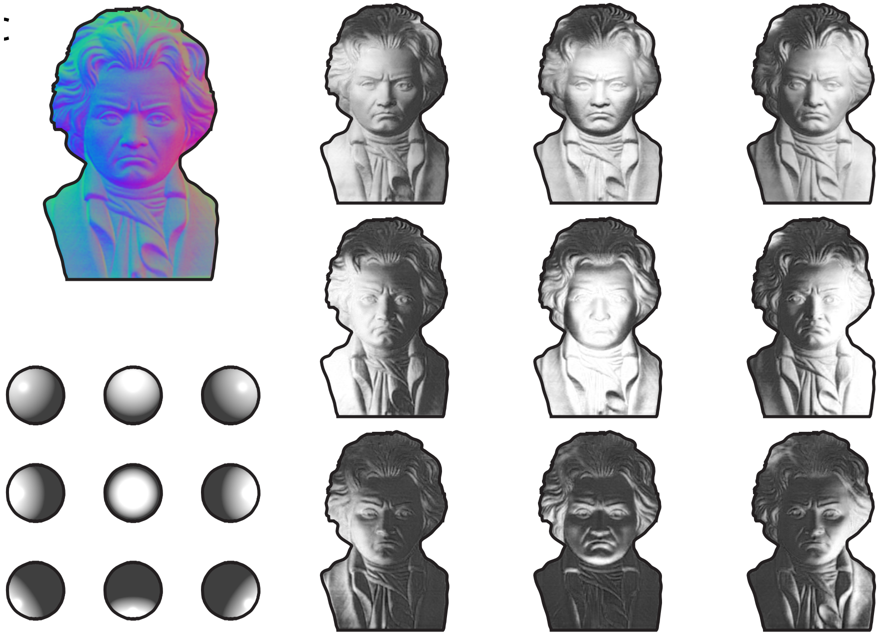


Causal intermediate stage visual representations of reflectance, shape, and lighting

Decomposing generic objects

Inputs

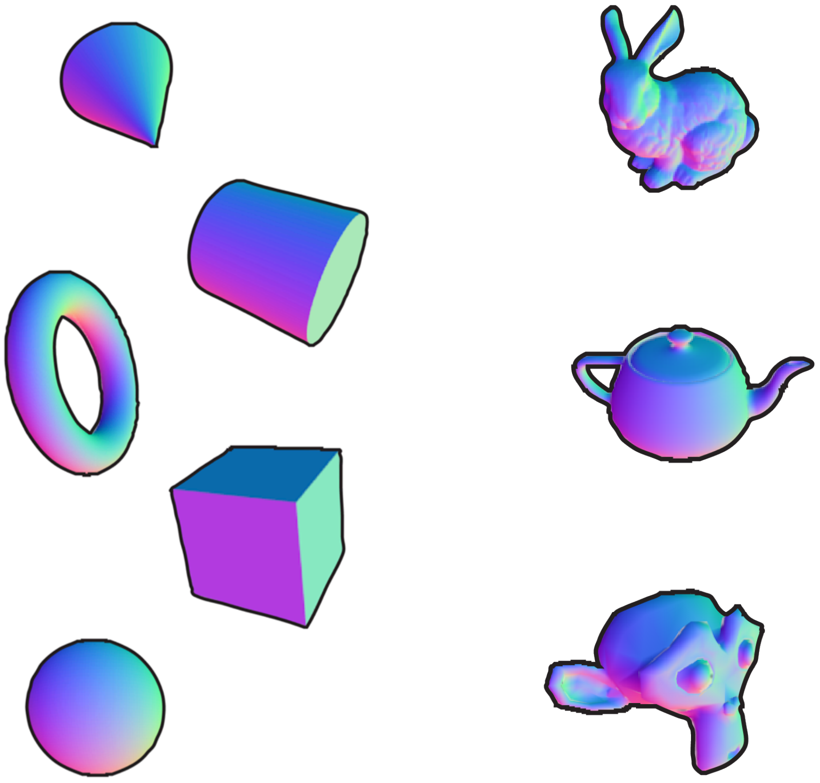
Shader Outputs



Causal intermediate stage visual representations of reflectance, shape, and lighting

Learn both the recognition model and the rendering function

Decomposing generic objects



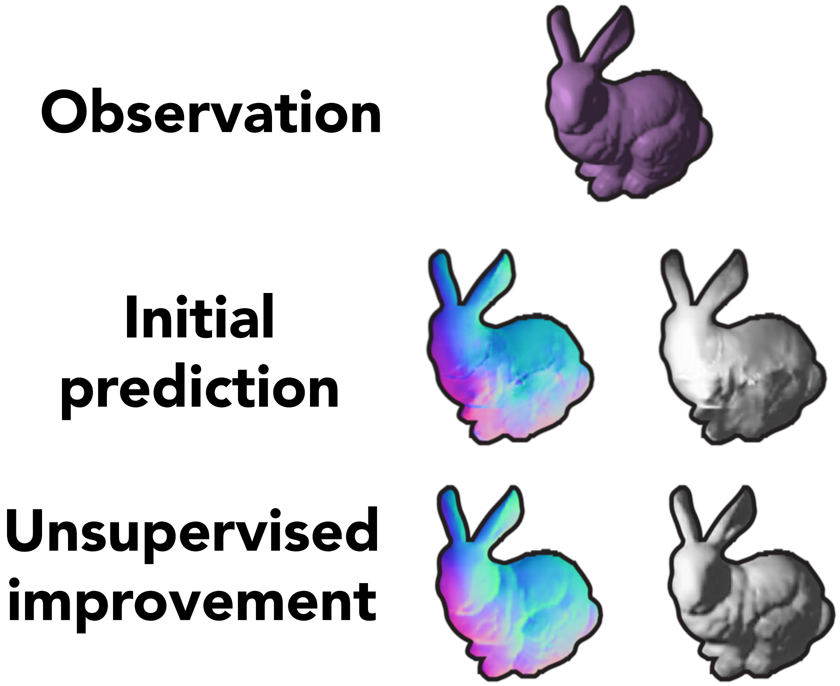
Train shapes **Test shapes**

Causal intermediate stage visual representations of reflectance, shape, and lighting

Learn both the recognition model and the rendering function

Representation learning driven by reconstruction error

Decomposing generic objects



Causal intermediate stage visual representations of reflectance, shape, and lighting

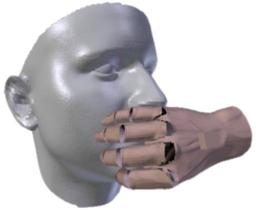
Learn both the recognition model and the rendering function

Representation learning driven by reconstruction error

Summary

- Causal and compositional models better predict human judgments

- Modeling causes allows for training-free crossmodal transfer



- Causal models can improve internal representations without supervision

Thank you

Ilker
Yildirim

Mario
Belledonne

Christian
Wallraven

Winrich
Freiwald

Joshua
Tenenbaum

Comparison Pipeline

Predict latents of both faces

Render study face with pose of test face

Occlude rendering and test image with mask

Compare in image (or latent) space