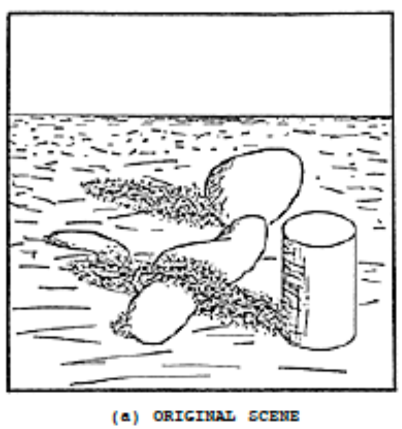
Intrinsic Images

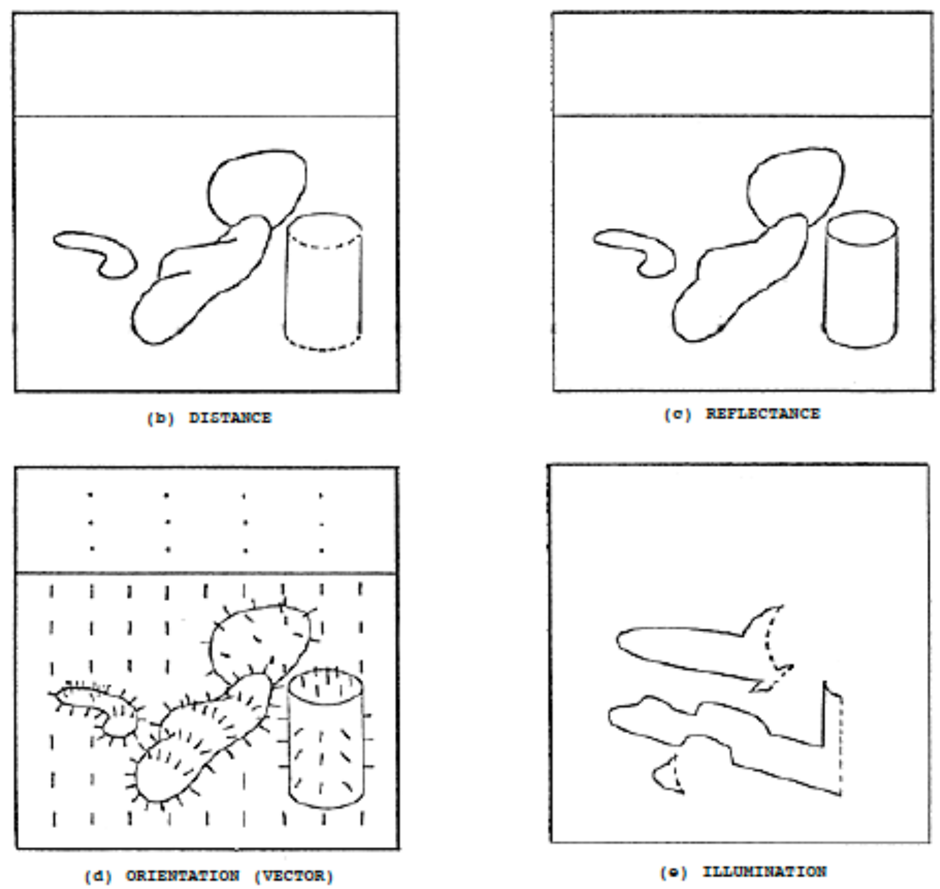
D.A. Forsyth, UIUC

Big technical points from the distant past

- Intrinsic images = maps of scene properties



Barrow+Tenenbaum, 1978



Intrinsics and extrinsics

- Intrinsic
  - Stuff that isn’t affected by lighting, weather, etc
  - wouldn’t change if you moved an object from image to image

- Extrinsic
  - Stuff that is
  - and would

Intrinsic images

- Intrinsic
    - shape, and affordances that follow
    - surface properties, and affordances that follow
    - volume properties, and affordances that follow

- Interesting because
  - What doesn’t change when
    - object moves from image to image?
    - light changes?

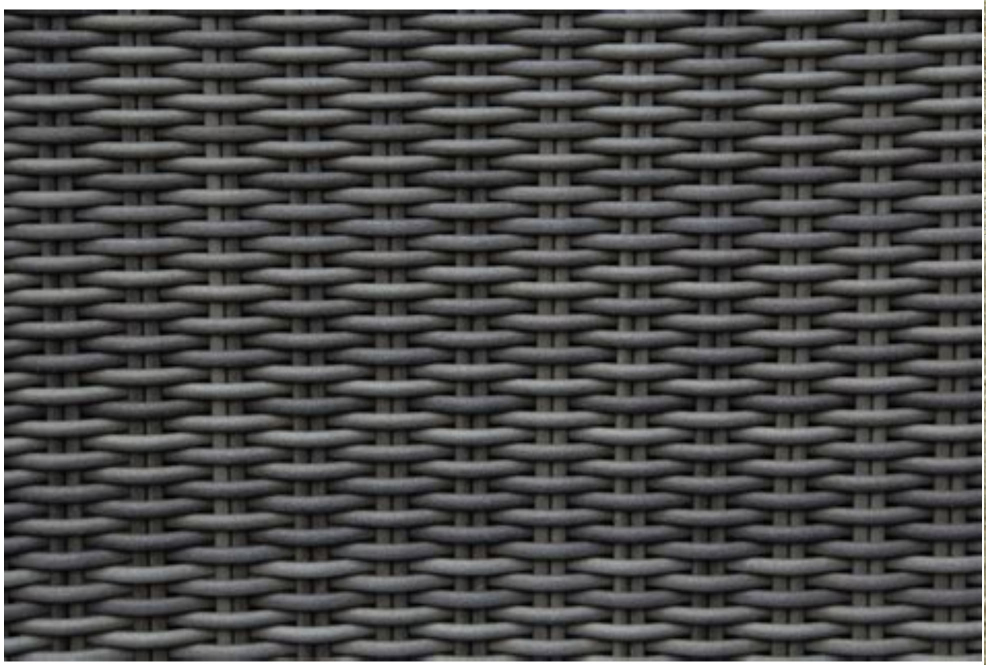
- Often dumbed down to albedo estimation

Possible Intrinsics

- Depth
- Normal
- Surface relief and material properties
  - and perhaps many of them
- Surface mechanical properties
- Surface glossiness
- Texture flow

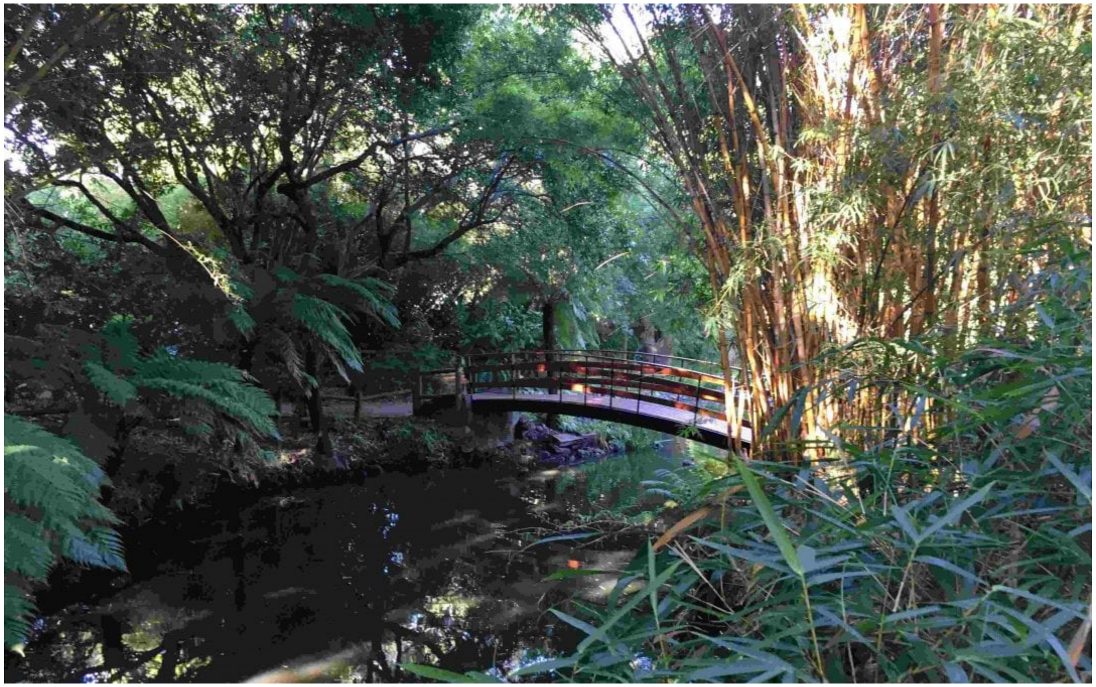
Relief - intrinsic, because small local shadows do not move with illumination (at least Koenderink+Van Doorn, 77)

Relief - intrinsic, because
small local shadows do not
move with illumination
(at least Koenderink+Van Doorn, 77)



Fur - intrinsic, because
small local shadows do not
move with illumination
(at least Koenderink+Van Doorn, 77)

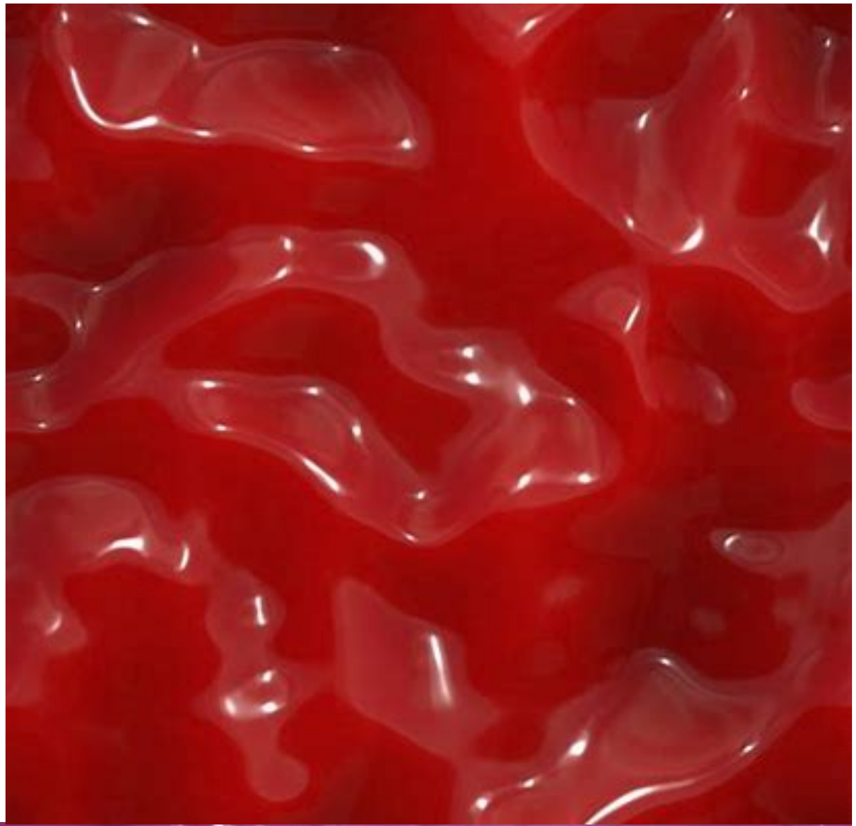
Relief - intrinsic (at least at this scale), because small local shadows do not move with illumination (at least Koenderink+Van Doorn, 77)

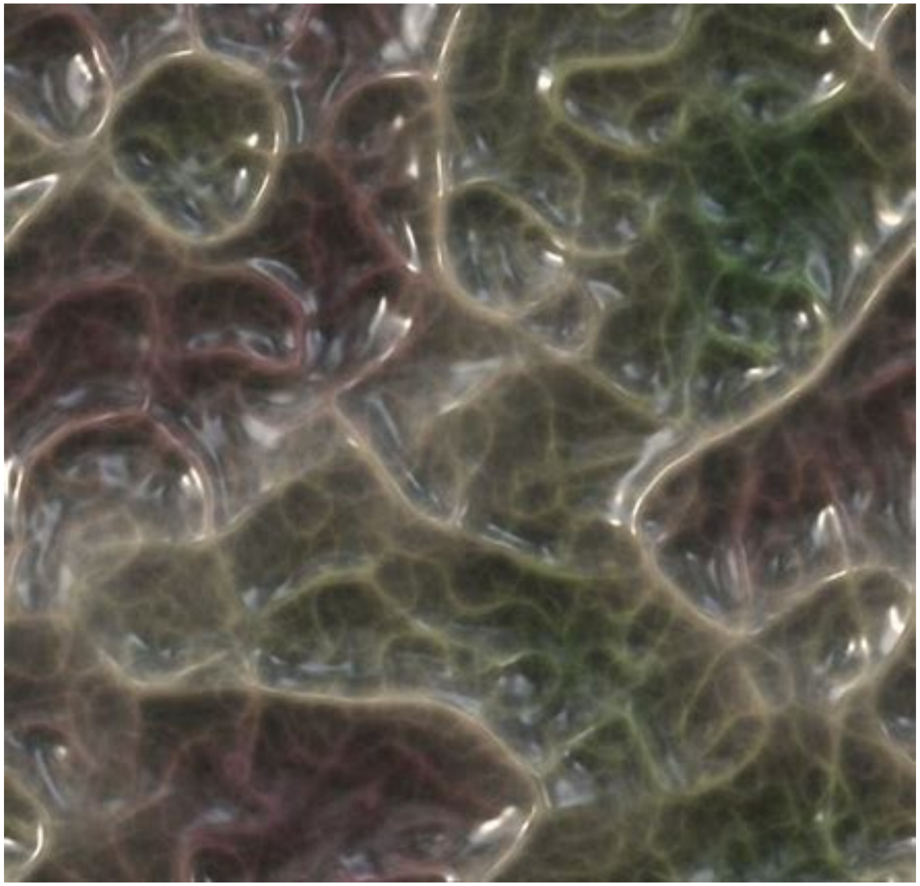


??? - intrinsic, because
mostly not a property of viewing
circumstances (?)

Iridescence
creating intrinsic gloss effects
intrinsic because the color effects will be
there for almost all illumination

??? - intrinsic, the specularities
move but are always there





??? - intrinsic, the specularities move but are always there

Other Possible Extrinsics

- Glossy reflected component
- Luminaires
- Lens flare
- Rain effects
- etc.

Gloss/specular - clearly extrinsic, when the light moves, this moves

Lens flares - clearly intrinsic, product of viewing circumstances

Luminaires -
extrinsic or intrinsic?
worth knowing about, anyhow

Rain - multiple extrinsic phenomena,
including smoothing, raindrops, loss of saturation,
glossy/wet surfaces, etc. etc.

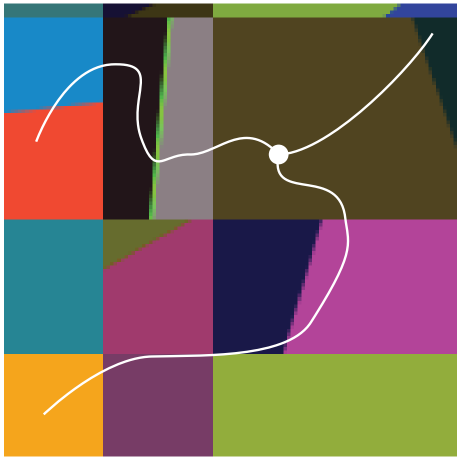
Why care about intrinsics…

- Different images of the same thing look different
  - under different lights
  - Consequences: classification problems; detection problems

From Flickr, webcam in Finland (SUNILA FI KAMERA)

Albedo/shading and Retinex

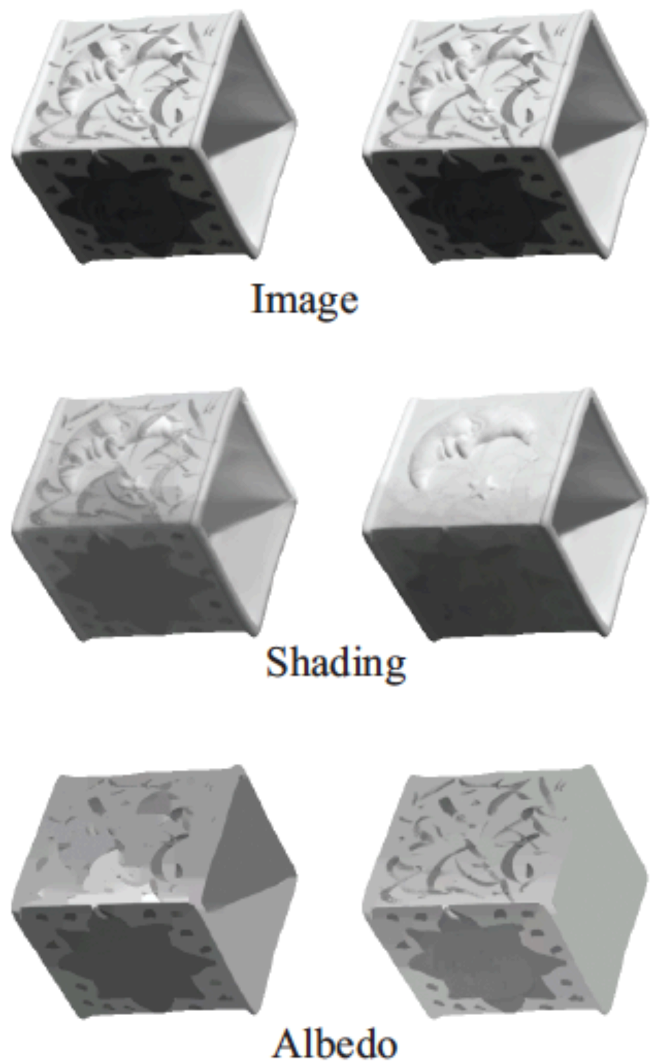
- Spatial reasoning, Land (59, 59, 77); Land +McCann 71:
  - Surface color changes either quickly or not at all
  - Light color changes slowly
- Retinex
  - big derivatives are albedo, small are shading;
    - DTI (Differentiate; threshold; integrate)
    - quite hard to know what Retinex does (Brainard+Wandell, 86)
    - large family of related algorithms inc Horn 73; Blake 85; Brelstaff+Blake 87; etc.

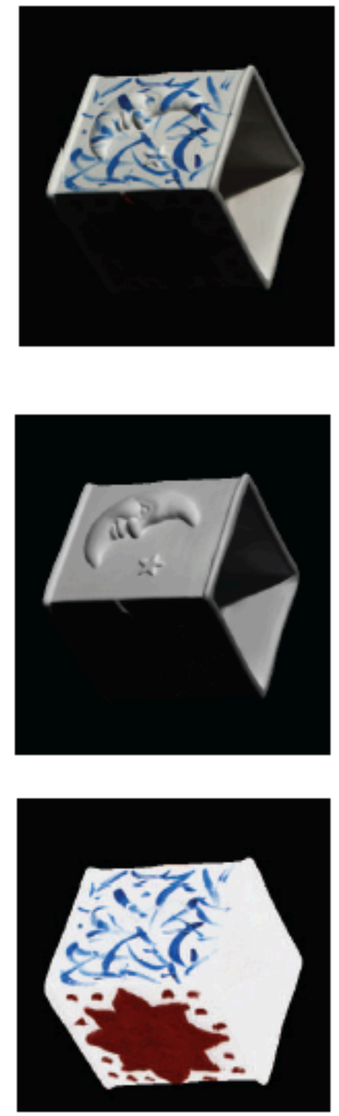


Retinex is really quite good

Implementation of Retinex
due to Kevin Karsch

Ground truth
images from dataset of Gosse et al. 09





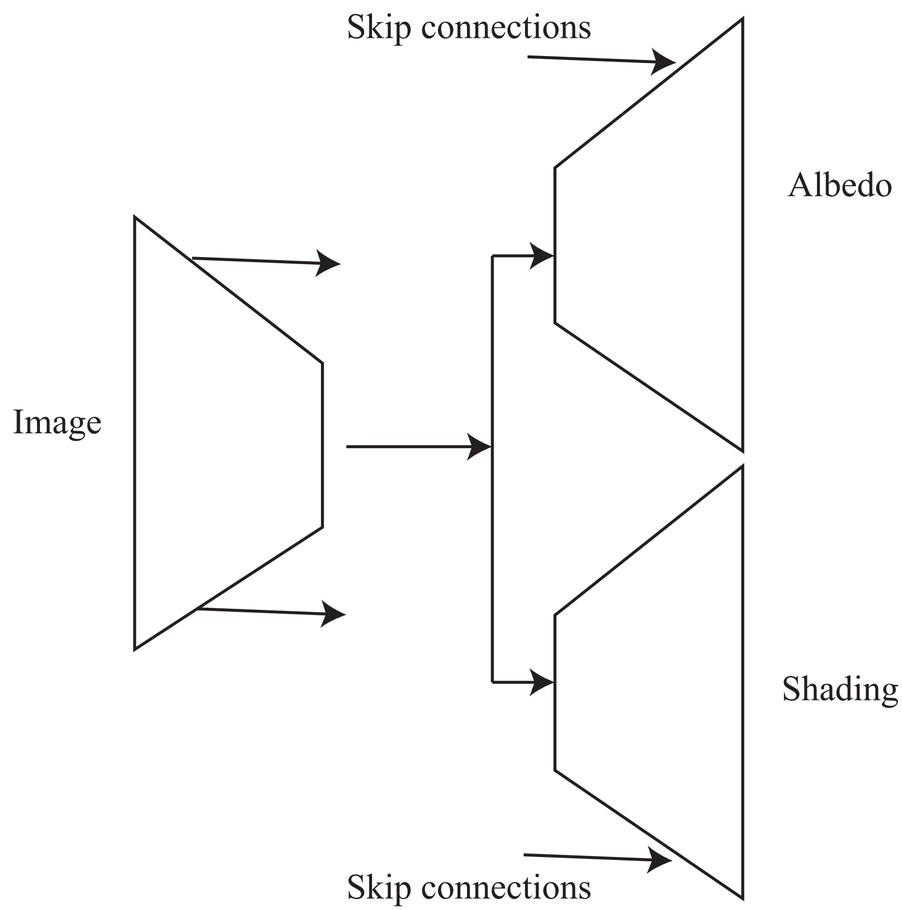
This gives an evaluation task

- WHDR=Weighted Human Disagreement Ratio
  - compute lightness from intrinsic image representation at points
  - predict
    - A lighter than B
    - B lighter than A
    - Lightness match
  - compute weighted estimate of accuracy
    - weights low where human judgements are uncertain, high otherwise
- There are issues, but allows evaluation
  - and competition

Replace DTI with network

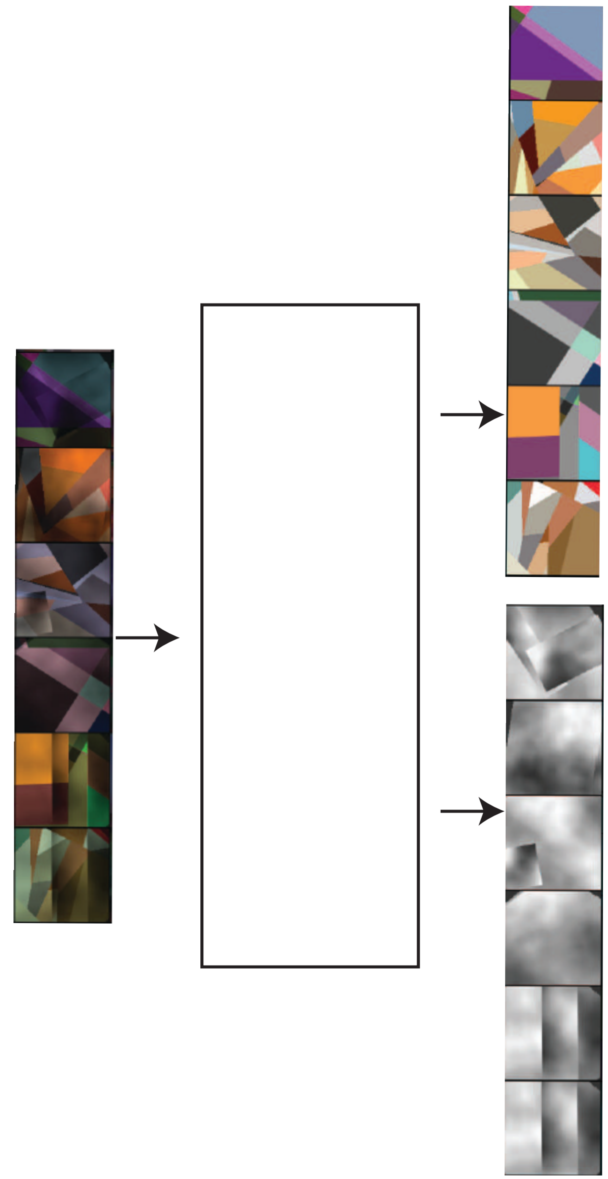
- Decomposition network
  - in goes image; out comes albedo, shading
- Train with
  - Fake images
    - multiply samples from spatial models
      - of albedo
      - of shading
    - Know the right answer, so loss is easy
  - Real images
    - use adversarial smoothing to ensure real albedos are “like” fake, etc.

One approach (local!)

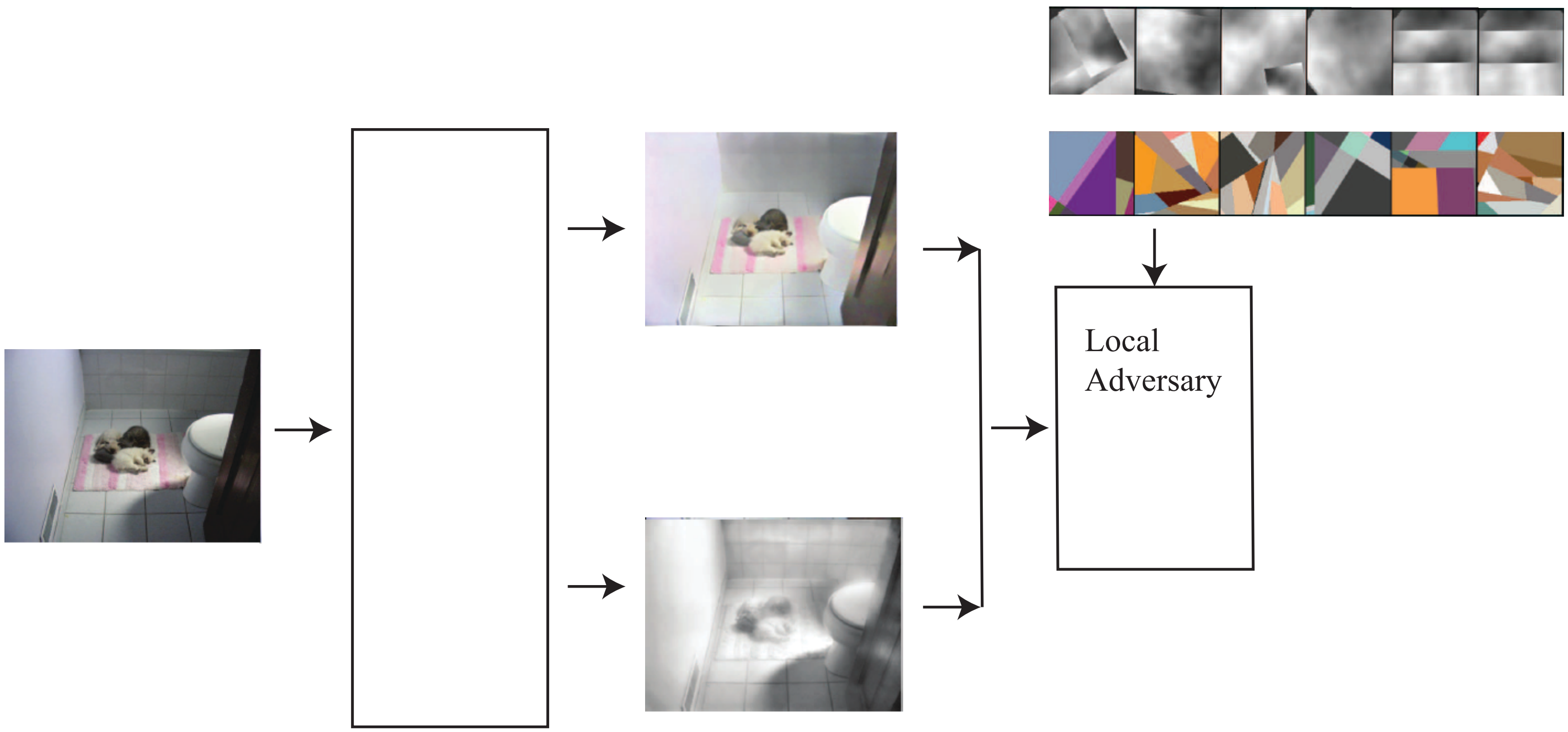


Training - I

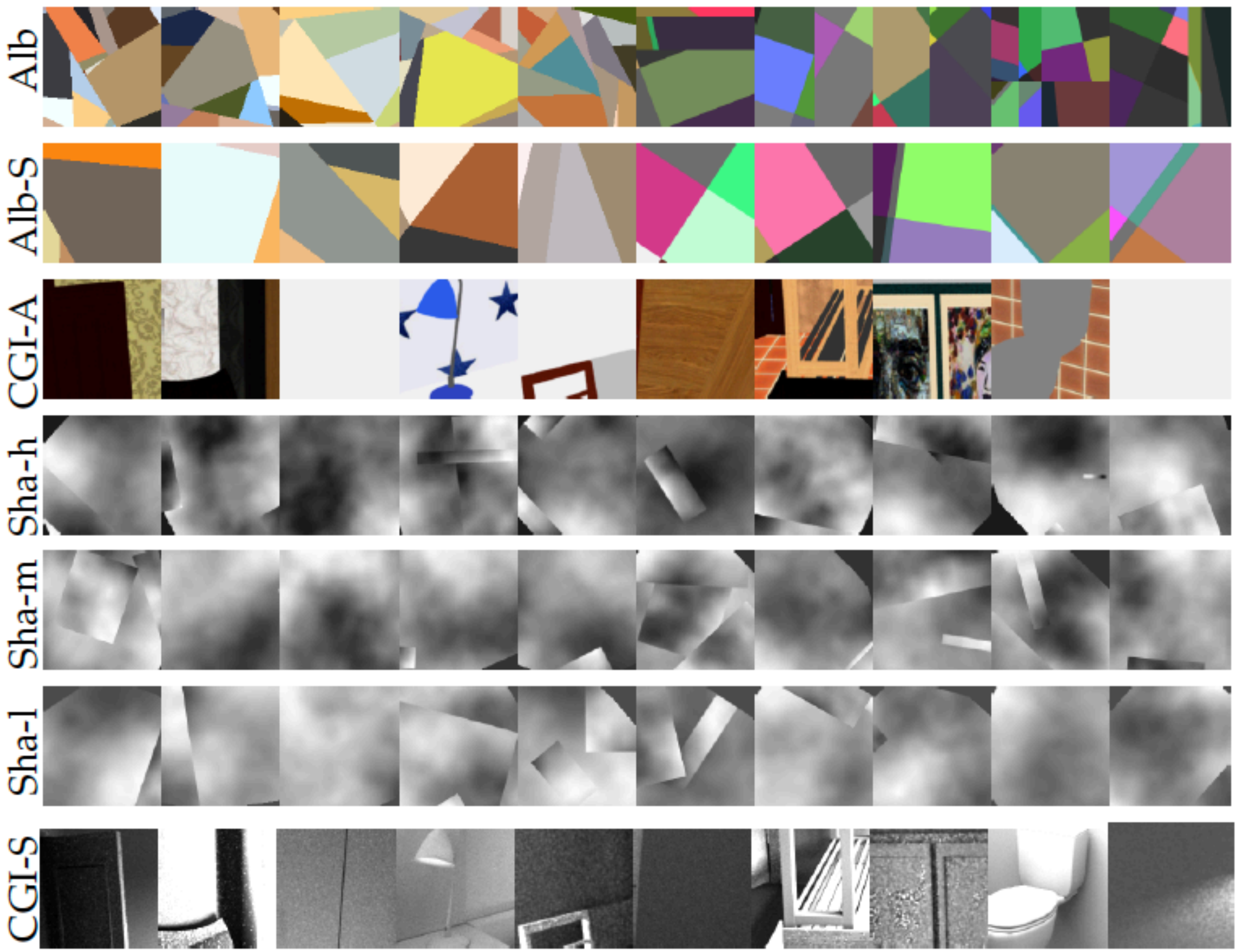
Our albedo paradigm uses a surface color model and a spatial model. The qualitative properties it is intended to capture are: albedoes are piecewise constant; the color distribution should reflect likely surface colors; there should be a profusion of edges with no strong orientation bias; there should be at least some vertices with degree greater than three. Surface color is modelled by drawing color samples uniformly and at random from the IIW training set. These must be adjusted for presumed illumination. We do so by assuming the range of illumination intensity is approximately the same as the range of lightnesses, and so dividing by the square root of intensity.



Training - II



Various options

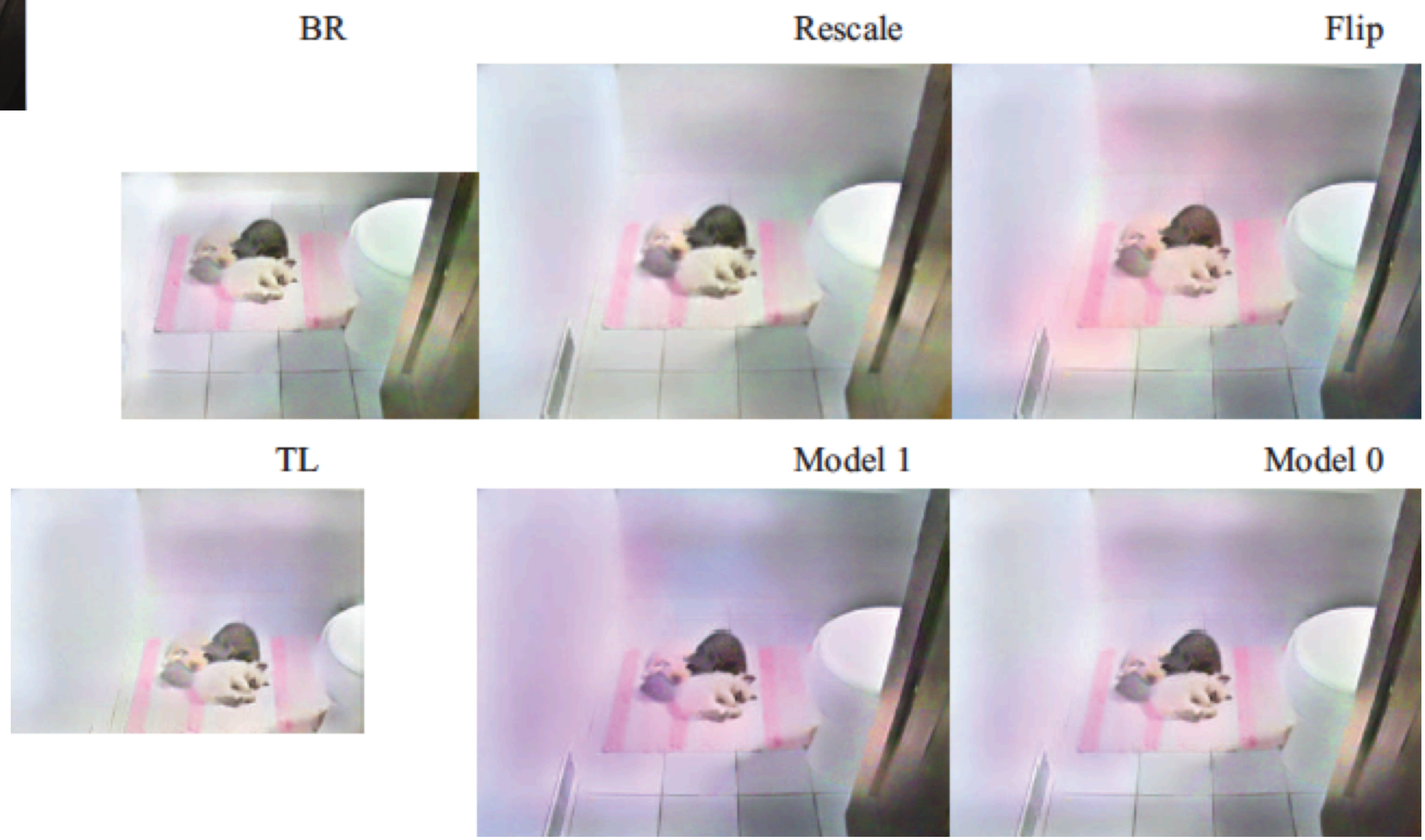


This story has a major problem

- Stopping training at different times yields different results
- Different crops of an image have different albedos
  - even at overlapping albedos

Nasty problem

- Translate, rotate, scale image
  - albedo for translated (etc) image should be translated albedo
  - shading for translated (etc) image should be translated shading
- But the network doesn't know that…



Averaging produces equivariance

In turn, this supplies a formal construction of an equivariant operation Ψ_{eq} out of any operation Ψ: we could simply average over G, to have

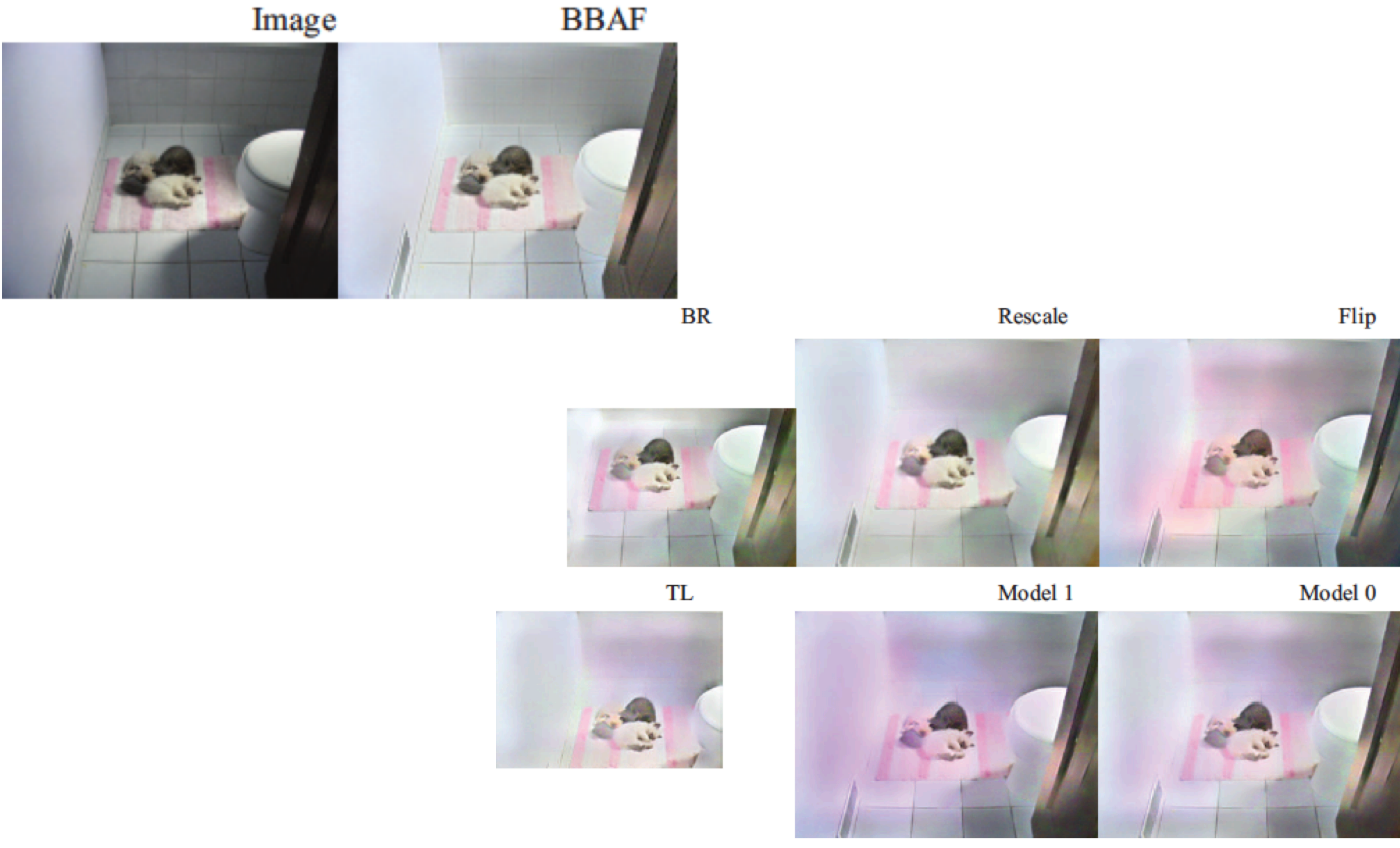
$$\Psi_{\mathrm{eq}}(f) = \left[\int_{g\in G} (g^{-1} \circ \Psi \circ g)(f) dg\right] / \left[\int_{g\in G} dg\right],$$

assuming that the integrals can be constructed, etc.

Imposing “equivariance” by averaging

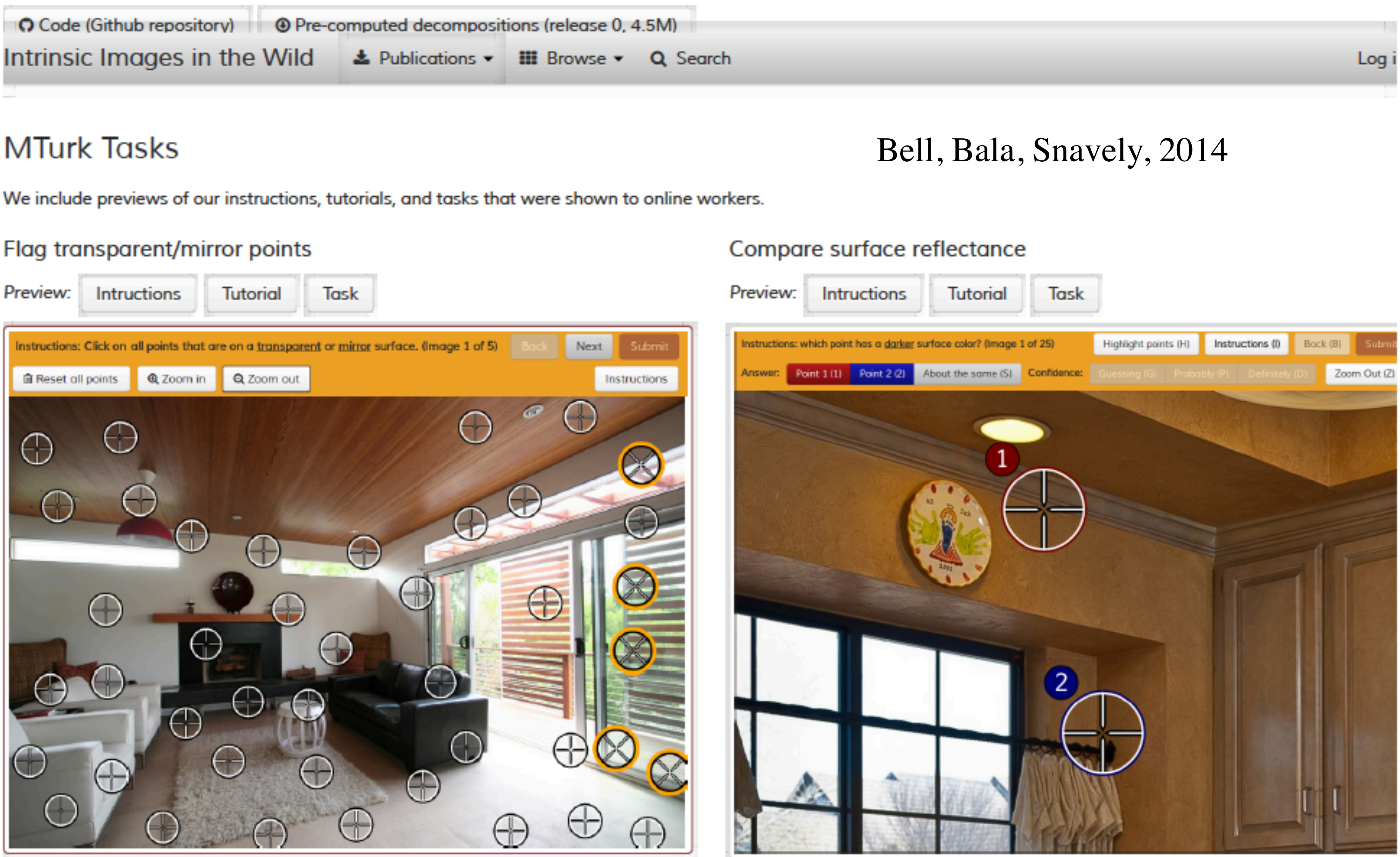
- We seek a class of equivariance property
- Adversarial smoothing:
  - Moving average of model coefficients
- Translation:
  - cover image with many, shifted, overlapping tiles
  - for each, recover albedo, shading
    - albedo at pixel is weighted average of all overlapping tiles
- Scale:
  - rescale image up, down
    - for each, recover albedo/shading using translation averaging
    - then rescale back
  - average results
- Rotation
  - average estimates from above over 8 flips (expensive)

Averaging very strongly suppresses error



Human judgements are easier

Bell, Bala, Snavely, 2014

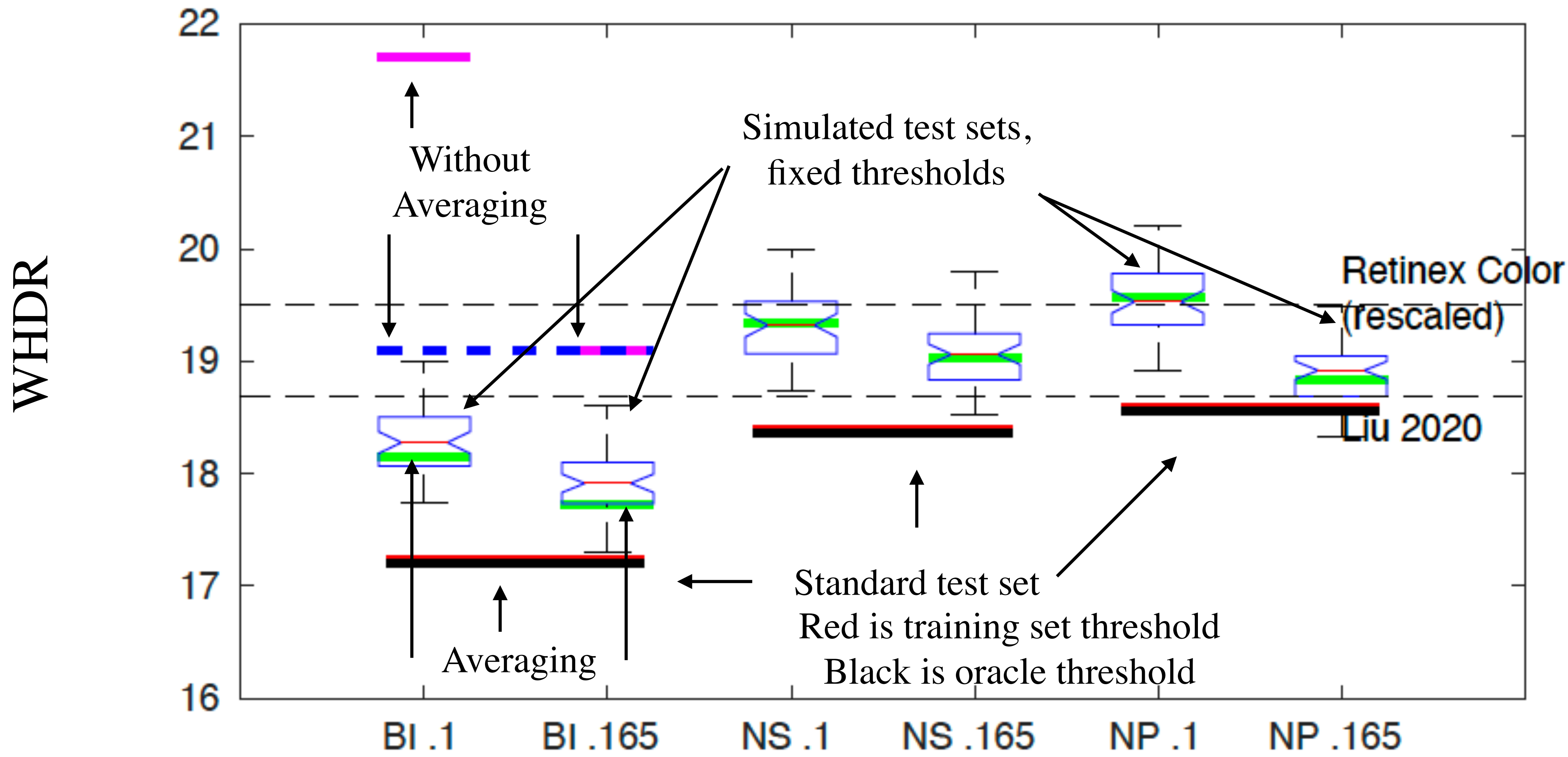


This gives an evaluation task

- WHDR=Weighted Human Disagreement Ratio
  - compute lightness from intrinsic image representation at points
  - predict
    - A lighter than B
    - B lighter than A
    - Lightness match
  - compute weighted estimate of accuracy
    - weights low where human judgements are uncertain, high otherwise
- There are issues (major ones!), but allows evaluation
  - and competition

Bell, Bala, Snavely, 2014

Averaging for equivariance is essential



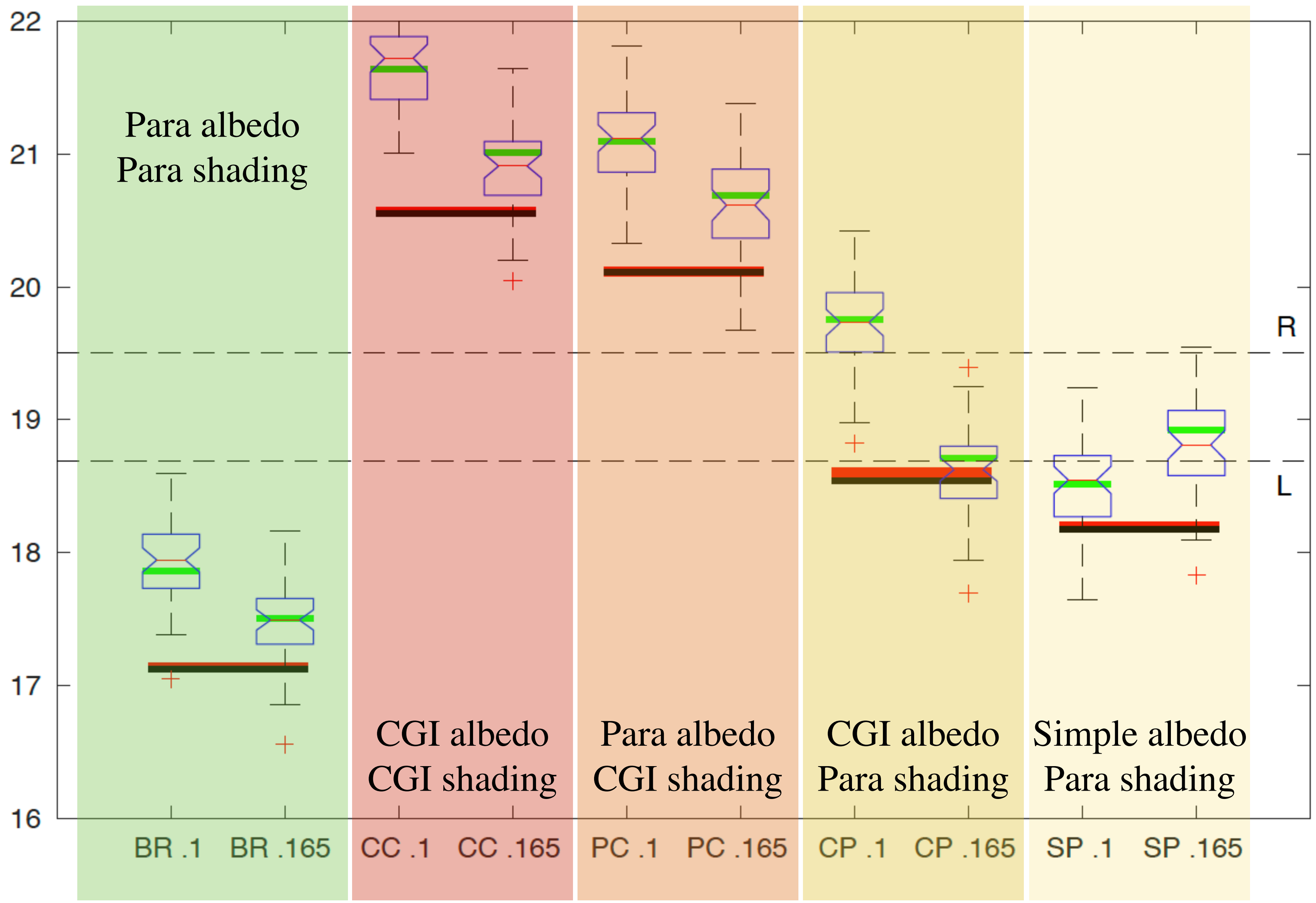
Strong model

no adversarial smoothing

no direct paradigms

DAF+Rock 21

CGI is a problem



Why is CGI not great?

CGI shading

CGI Shading noise
CGI shading is “simple”

Para shading

CGI albedo

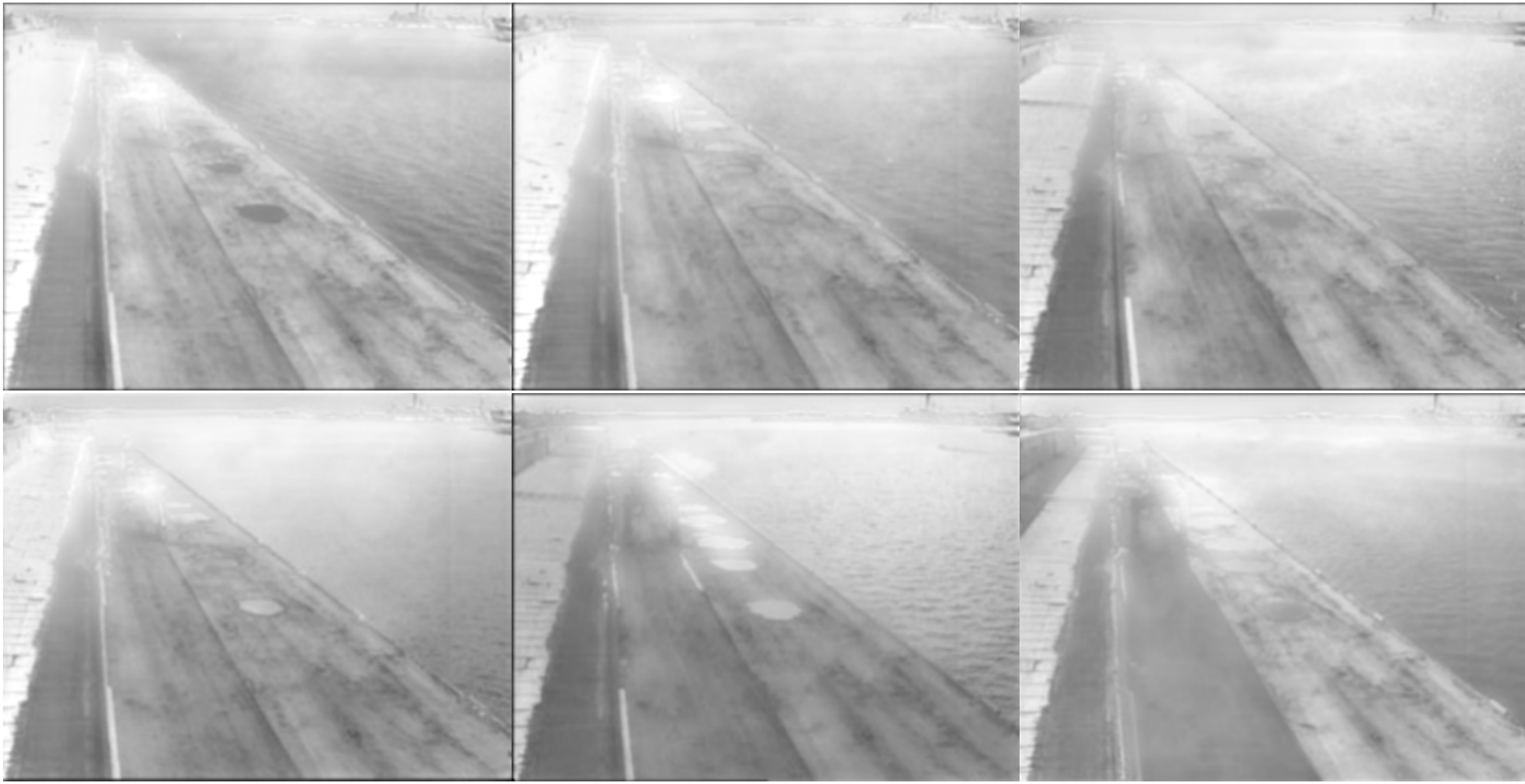
CGI albedo is “simple”

Para albedo

Paradigms are aggressive summaries of real problems

Paradigms pack pixel problems prodigiously

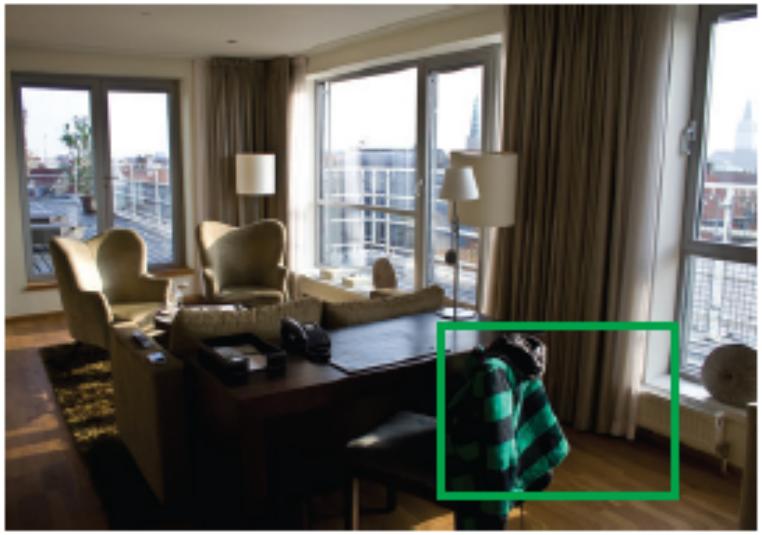
Finnish webcam



Actually, there is a snake in this garden

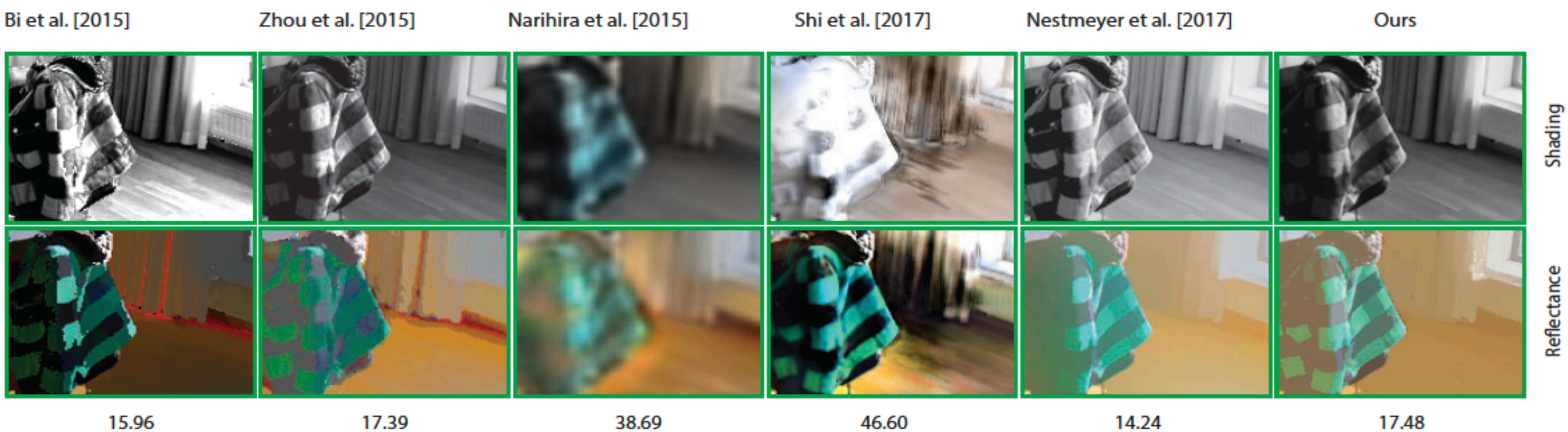
Annoying properties of current models

- Weird albedos
  - likely to do with WHDR evaluation
- Indecisiveness
  - Deep and poorly understood
- Poor behavior on multi-image datasets



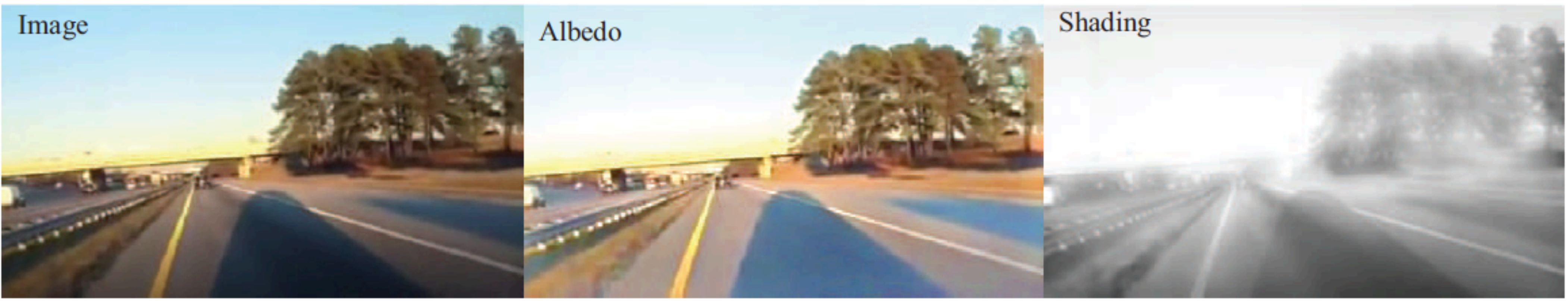
OFFICE

Indecisiveness



From Bi et al 18

Indecisiveness remains (aargh!)



Likely a problem with shading…

- Albedo recovery should be invariant to lighting changes
- Q:
  - Is it?

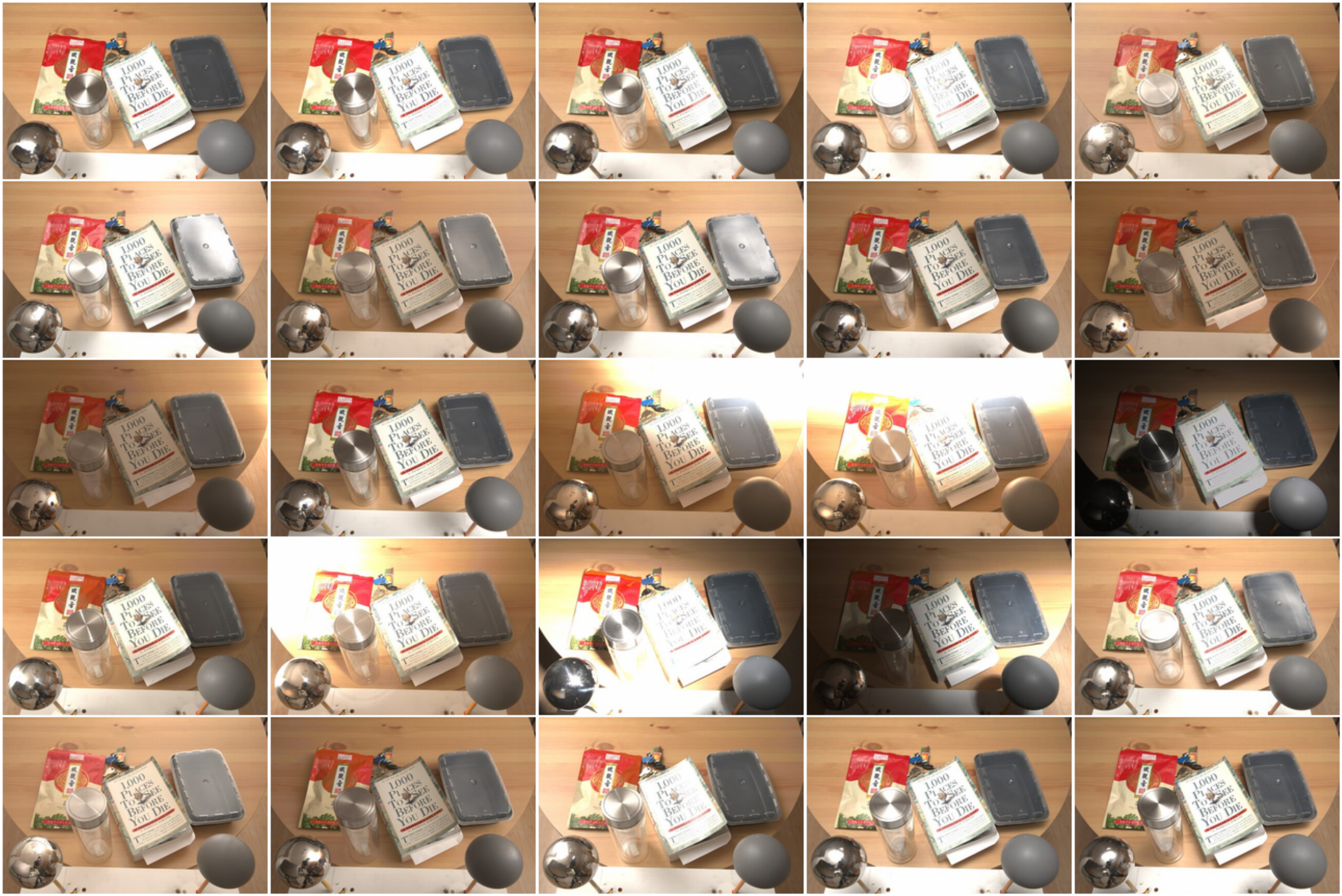
Neat Dataset: MIT Multi-Image

A Dataset of Multi-Illumination Images in the Wild

Lukas Murmann[*1] Michael Gharbi[1 2] Miika Aittala[1] Fredo Durand[1]
[1]MIT CSAIL [2]Adobe Research [*]lmurmann@mit.edu

Collections of images under a single, uncontrolled illumination have enabled the rapid advancement of core computer vision tasks like classification, detection, and segmentation. But even with modern learning techniques, many inverse problems involving lighting and material understanding remain too severely ill-posed to be solved with single-illumination datasets. To fill this gap, we introduce a new multi-illumination dataset of more than 1000 real scenes, each captured under 25 lighting conditions.

https://projects.csail.mit.edu/illumination/



https://projects.csail.mit.edu/illumination/


1,000 PLACES TO SEE BEFORE YOU DIE

Likely a problem with shading…

- Albedo recovery should be invariant to lighting changes
- Q:
  - Is it?

NOT EVEN SLIGHTLY

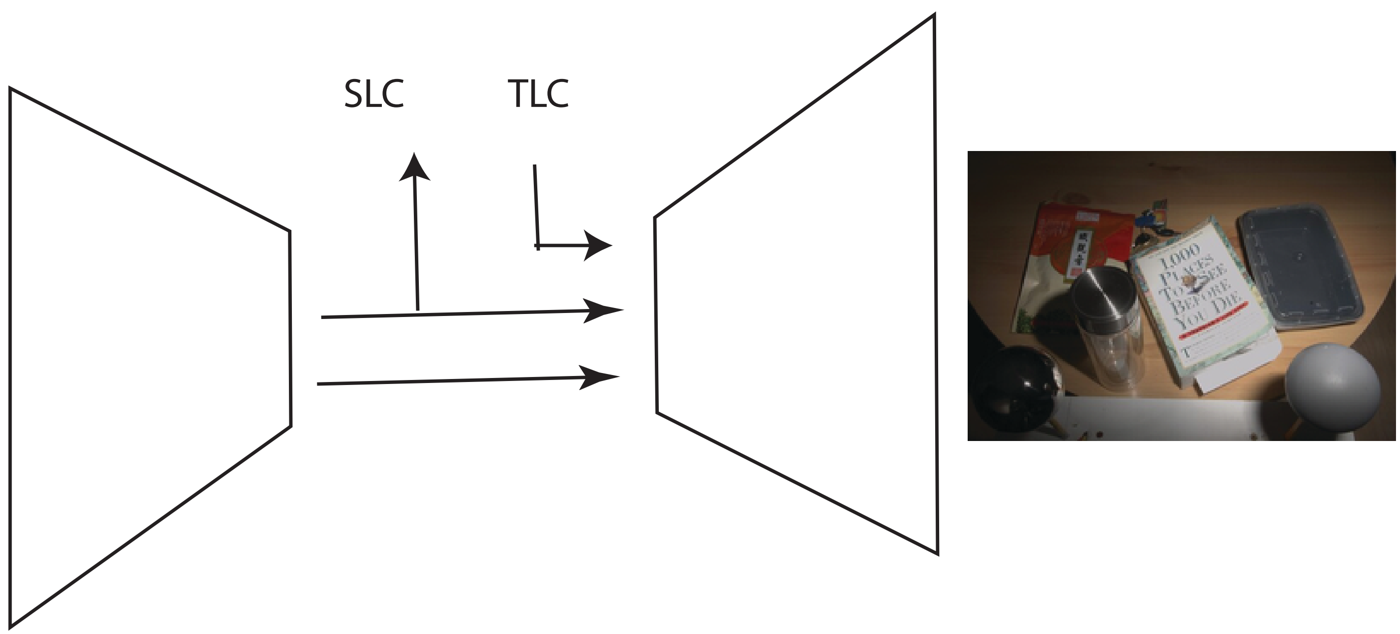
- Q: What to do?
  - Averaging?
  - Augmentation?

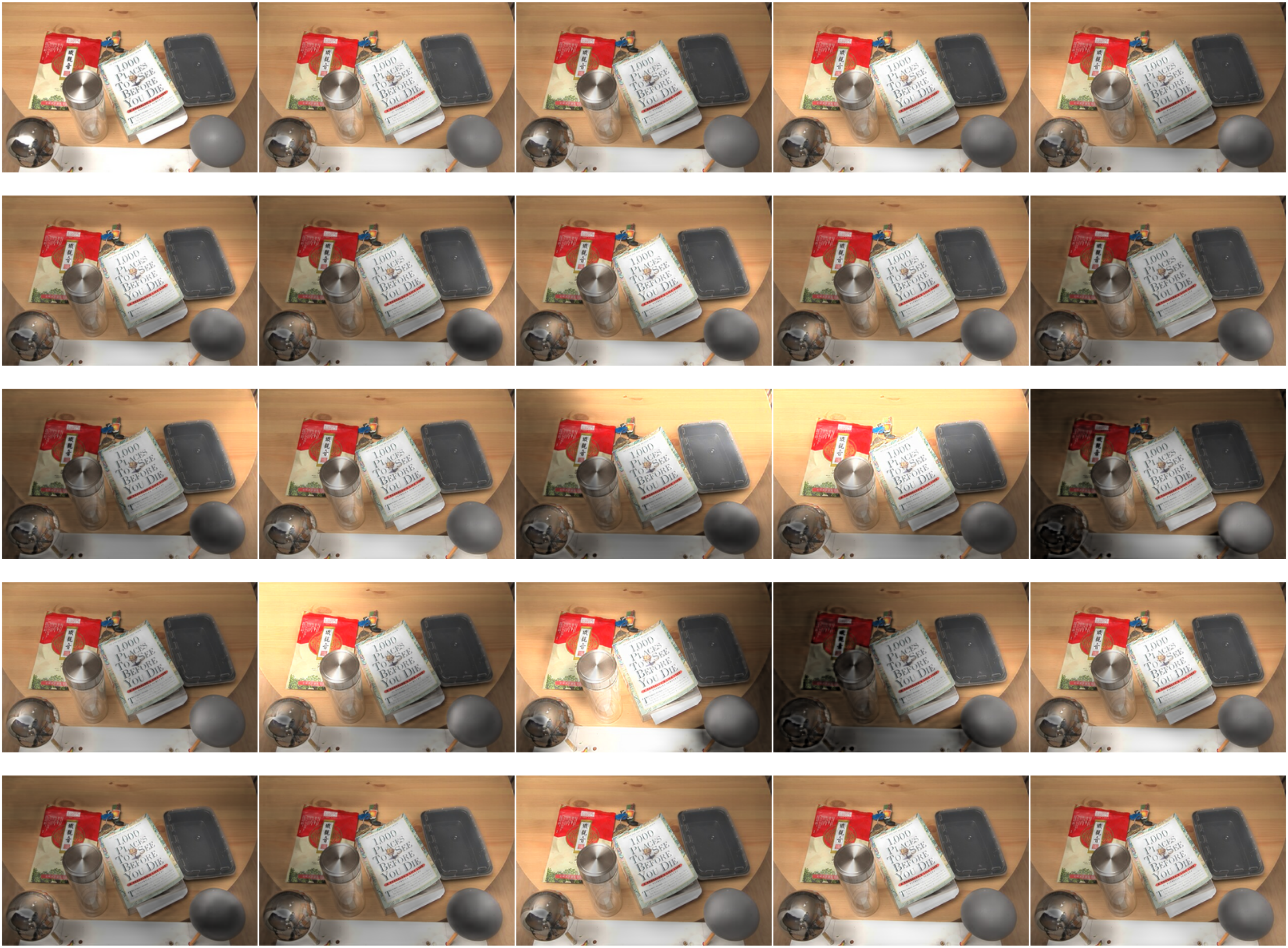
HOW?

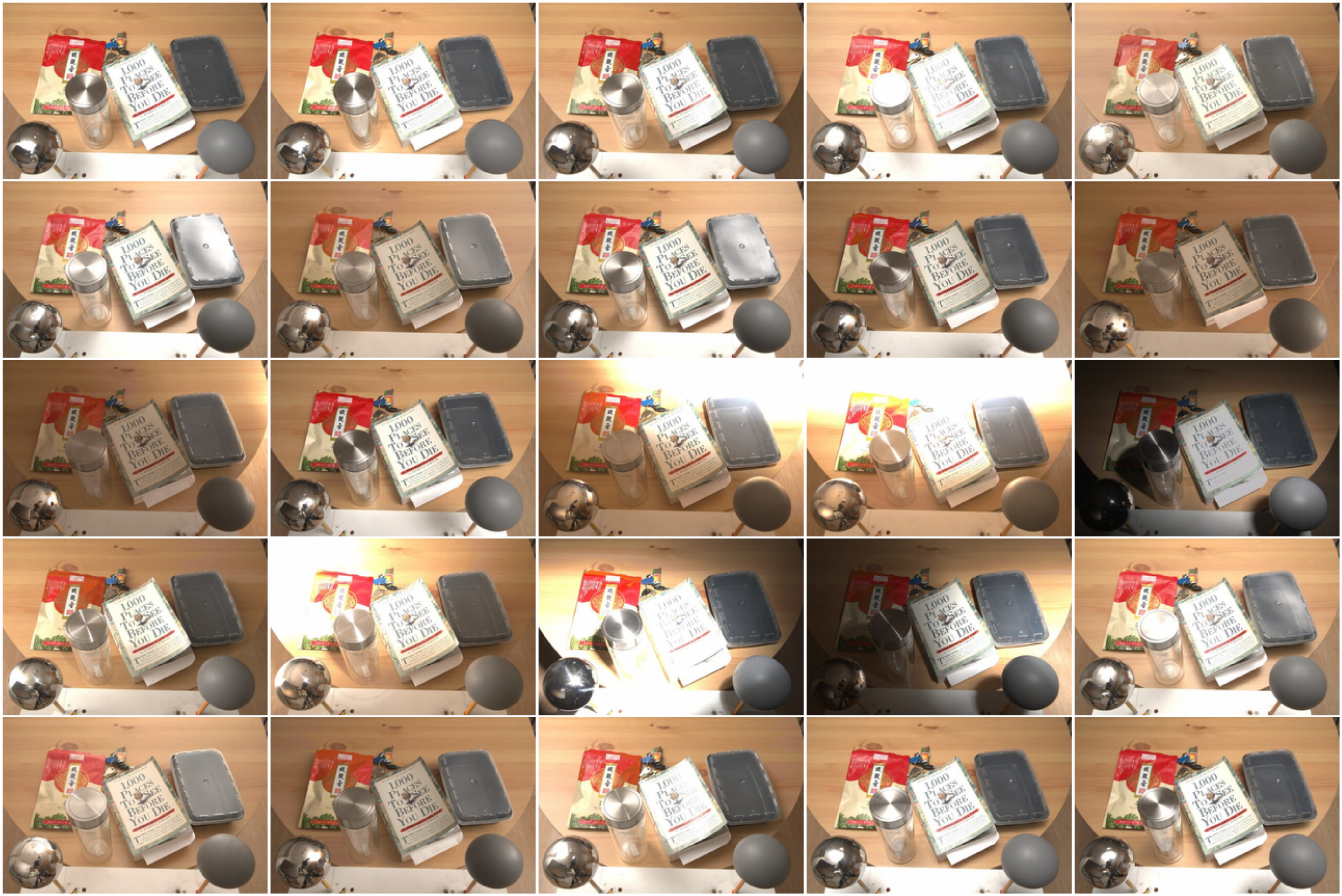
Relighting to Suppress Variance

- MIT dataset has a special structure
  - illlumination is known and controlled (25 illums per scene)
    - image_ij = scene_i x illum_j
      - where j'th illum is the same across scenes
- This makes it "easy" to build a relighter
  - illumination rep. w/ code (SLC - source lighting code; TLC - target etc)
  - train w/ L1L2 loss and adversary

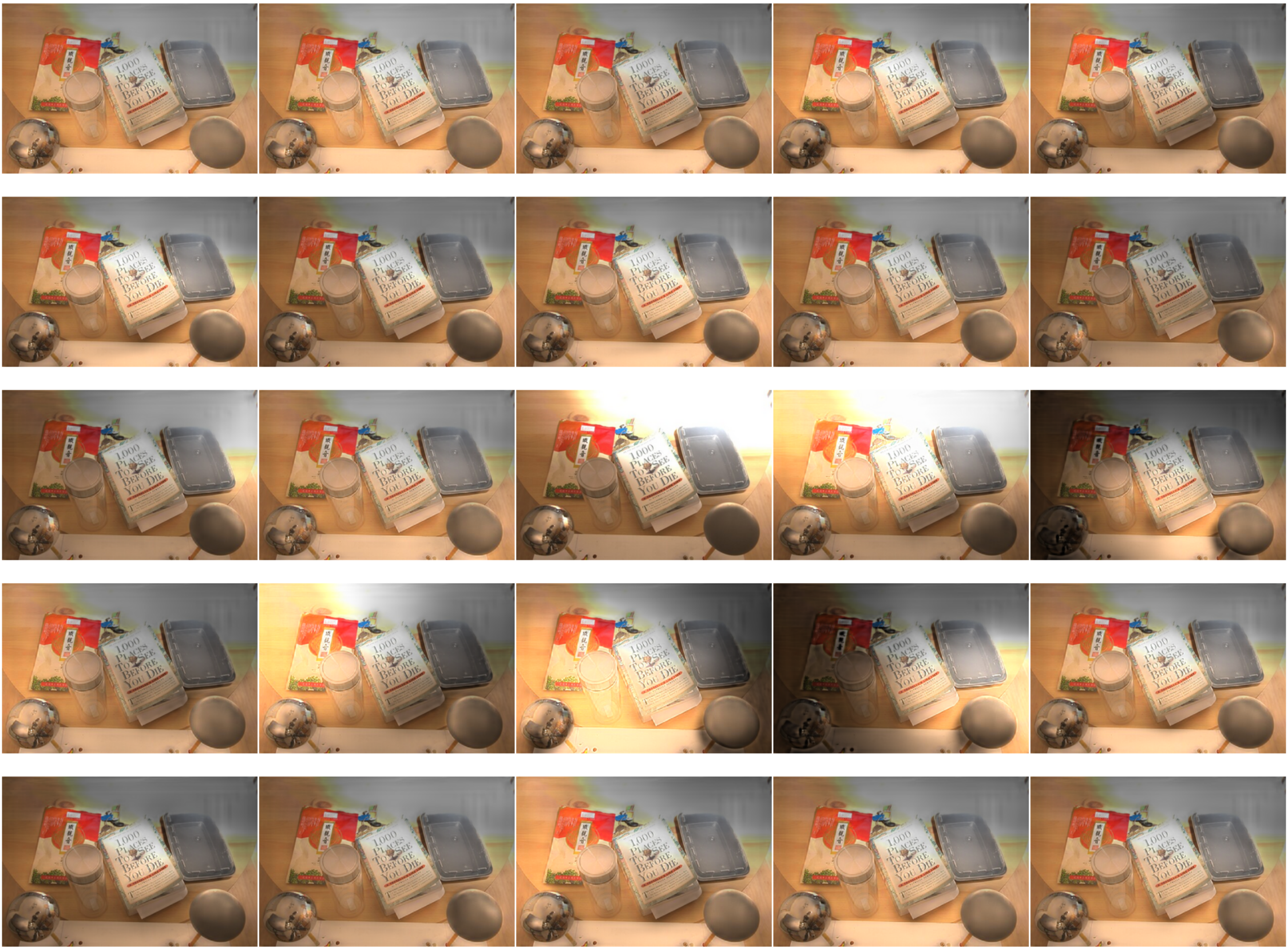
(optionally, depth+normal)

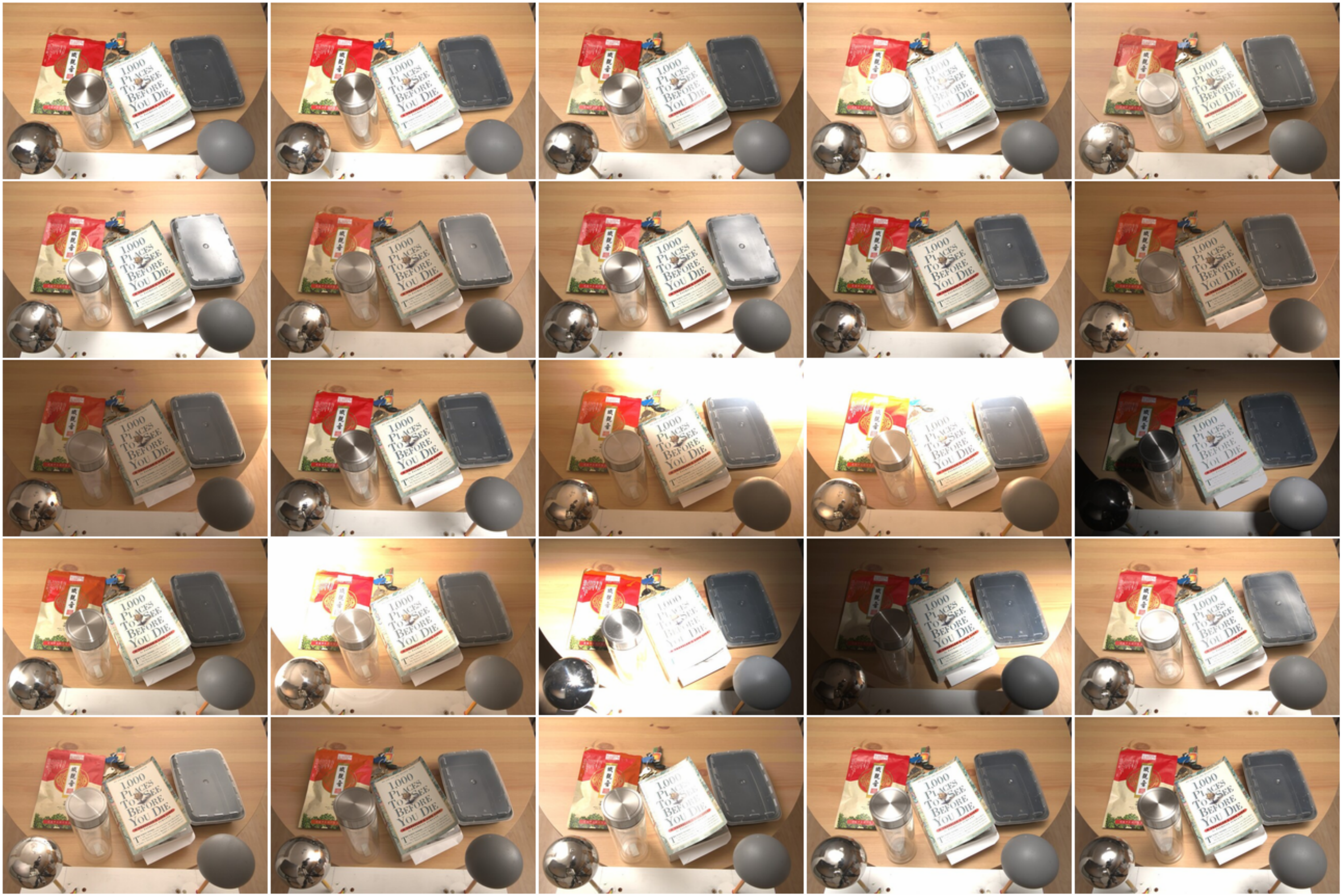



1,000
PLACES
TO SEE
BEFORE
YOU DIE



https://projects.csail.mit.edu/illumination/


1,000 PLACES TO SEE BEFORE YOU DIE



https://projects.csail.mit.edu/illumination/

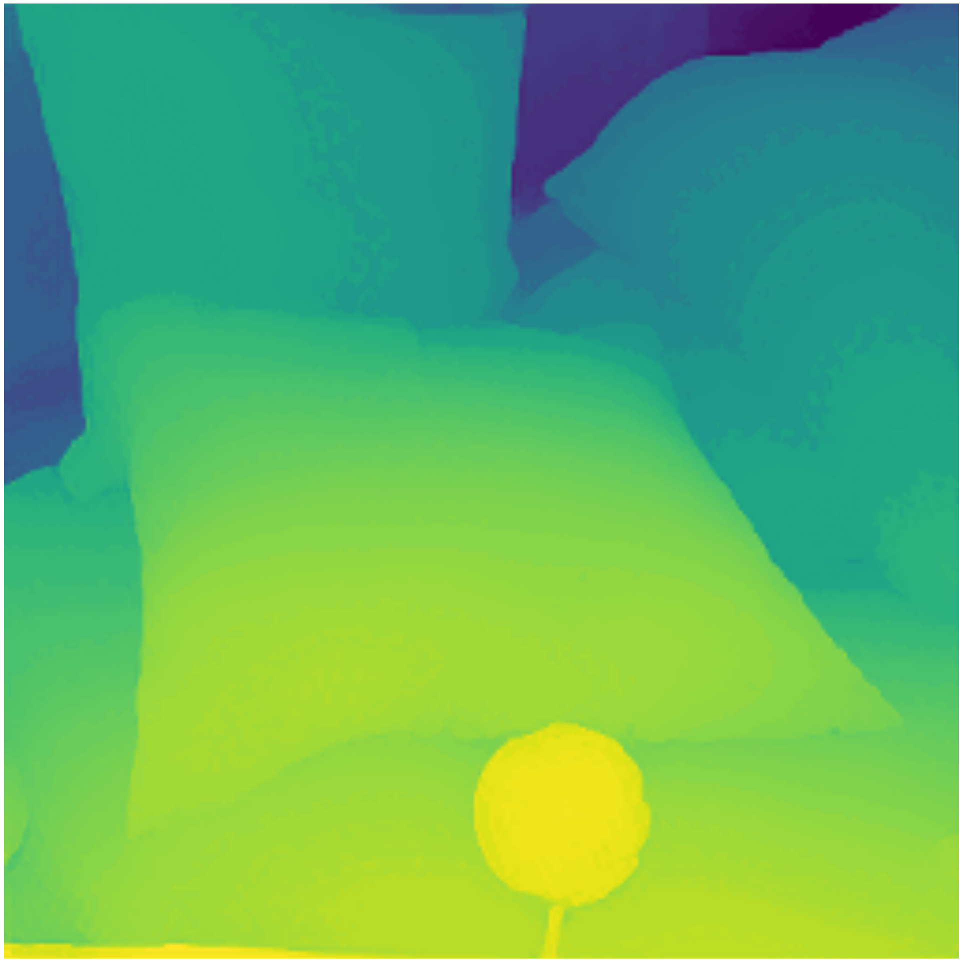
NLA

VCA (Image+depth+normal)

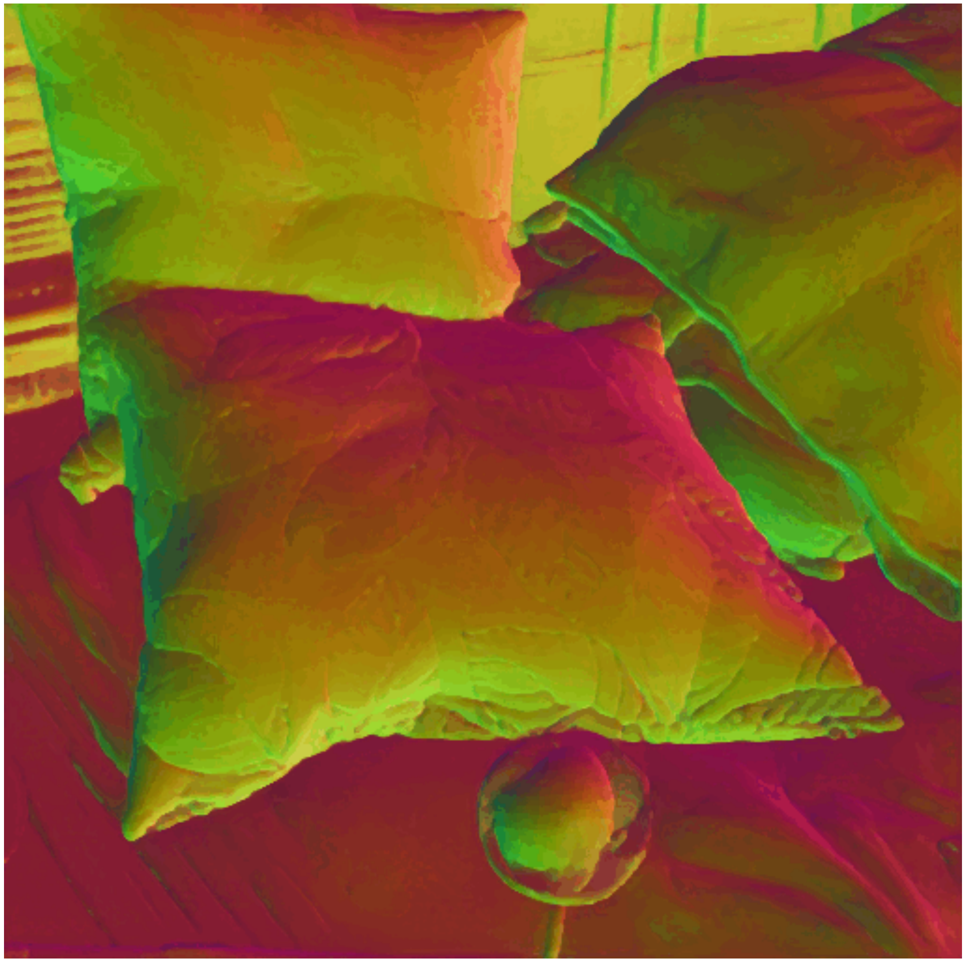
Litany of questions

- Is there an issue for depth/normal?
- Are effects big enough to care about?
- How well does averaging suppress effects of illumination?
  - compare NLX with VCX (X={A, N, D})
  - 25 estimates per scene, one per illumination
  - Look at var(VCX)/var(NLX) using appropriate metrics
- Is it cheating?
  - compare means
- Are some relighting models better than others?
- Are some scenes harder than others?

Depth (omnimap, current best depth est)



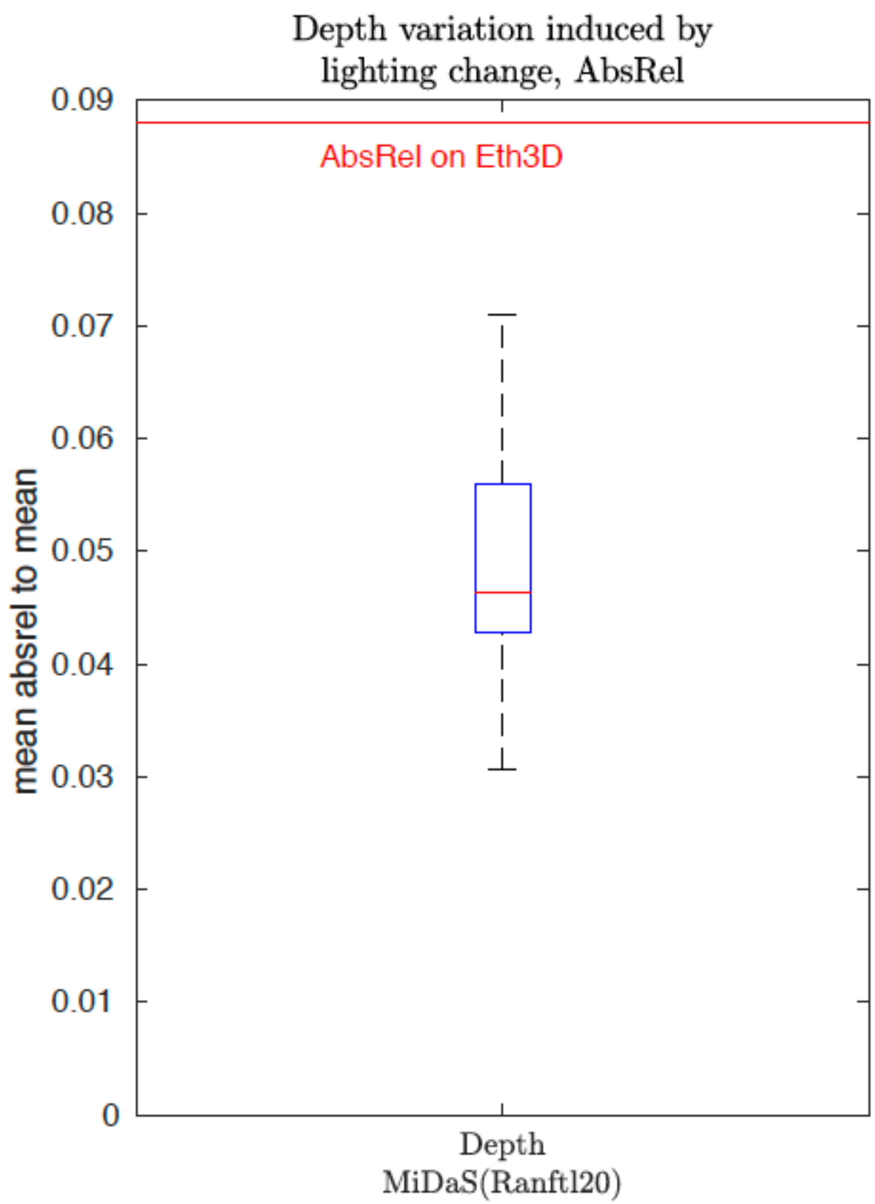
Normal (omnimap, current best normal est)

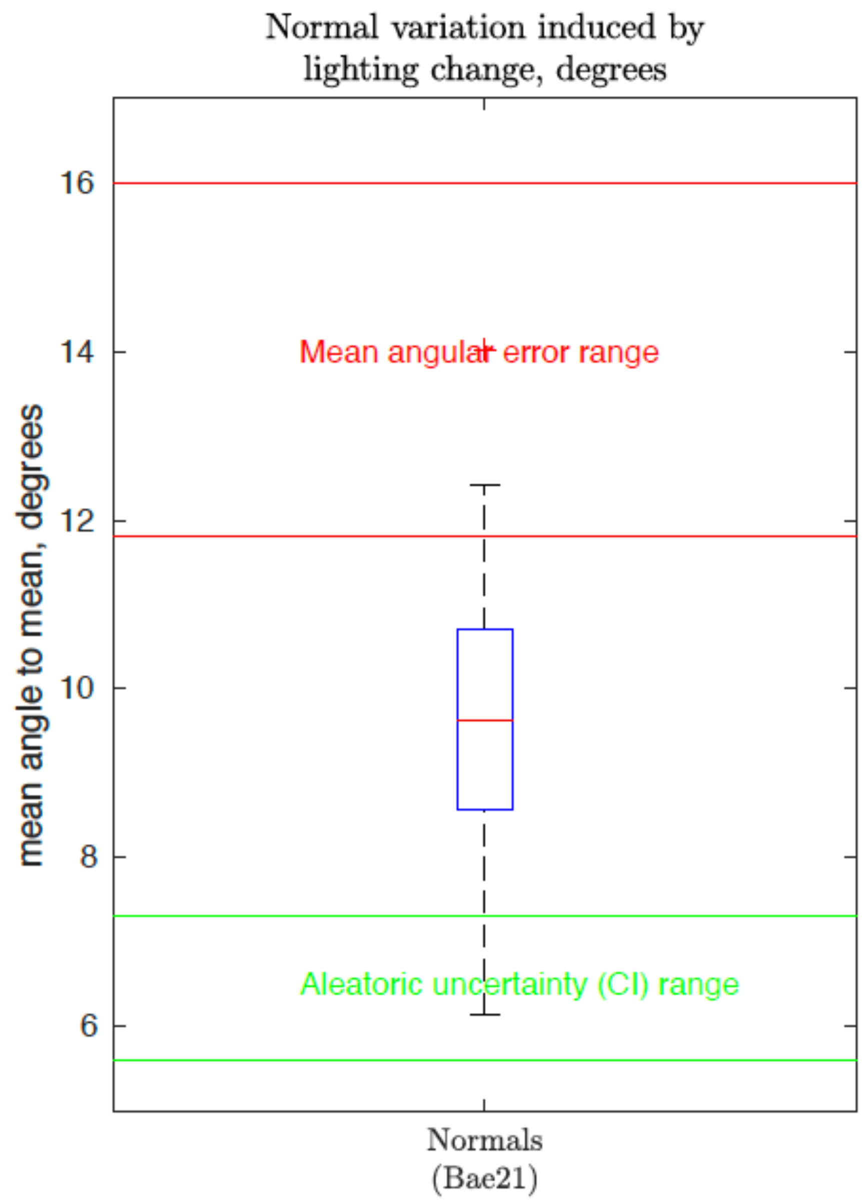


Litany of questions

- Is there an issue for depth/normal?
- Is it big enough to care about?
- How well does averaging suppress effects of illumination?
  - compare NLX with VCX (X={A, N, D})
  - 25 estimates per scene, one per illumination
  - Look at var(VCX)/var(NLX) using appropriate metrics
- Is it cheating?
  - compare means
- Are some relighting models better than others?
- Are some scenes harder than others?

Is there an issue for depth/normal?

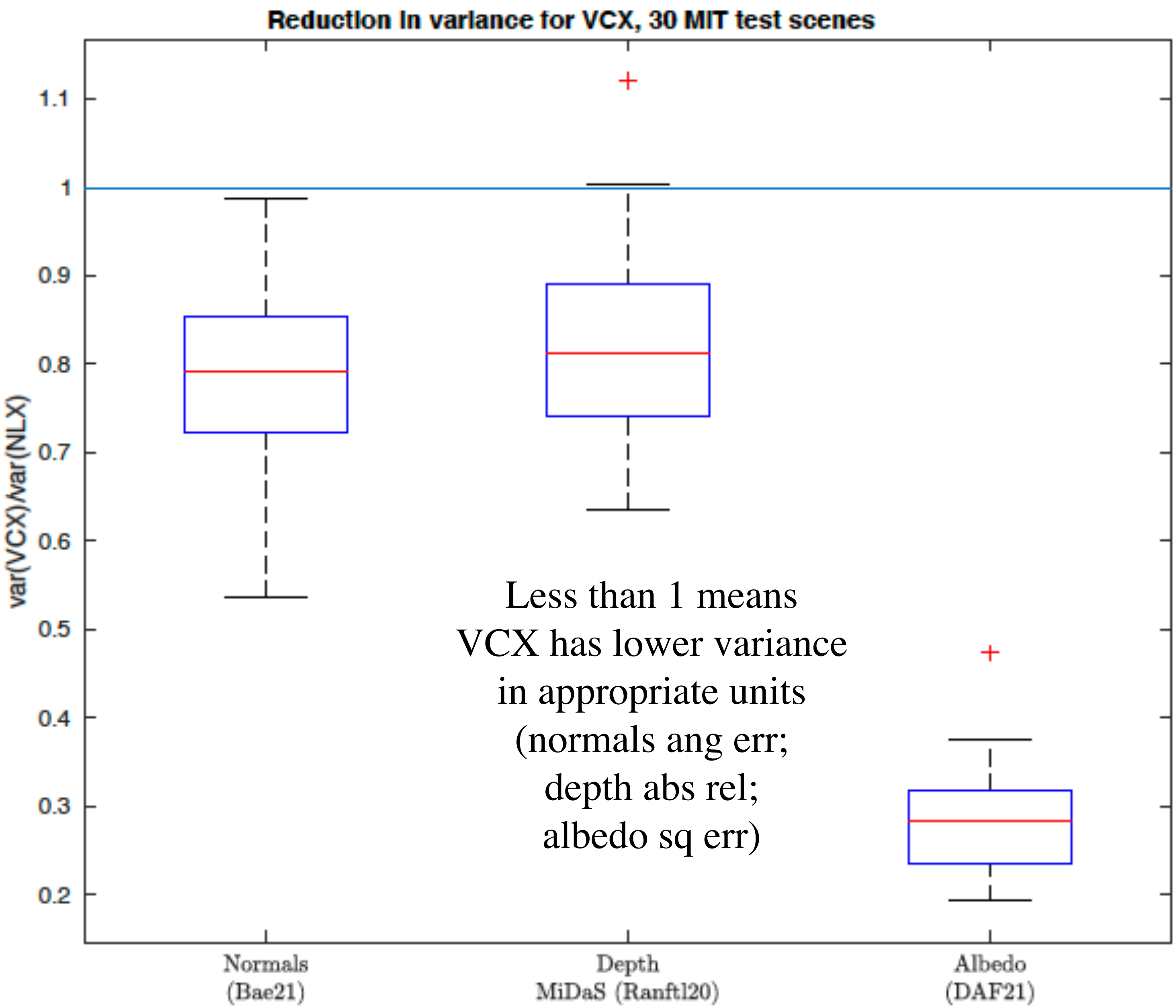




Litany of questions

- Is there an issue for depth/normal?
- Is it big enough to care about?
- How well does averaging suppress effects of illumination?
    - compare NLX with VCX (X={A, N, D})
    - 25 estimates per scene, one per illumination
    - Look at var(VCX)/var(NLX) using appropriate metrics
- Is it cheating?
    - compare means
- Are some relighting models better than others?
- Are some scenes harder than others?

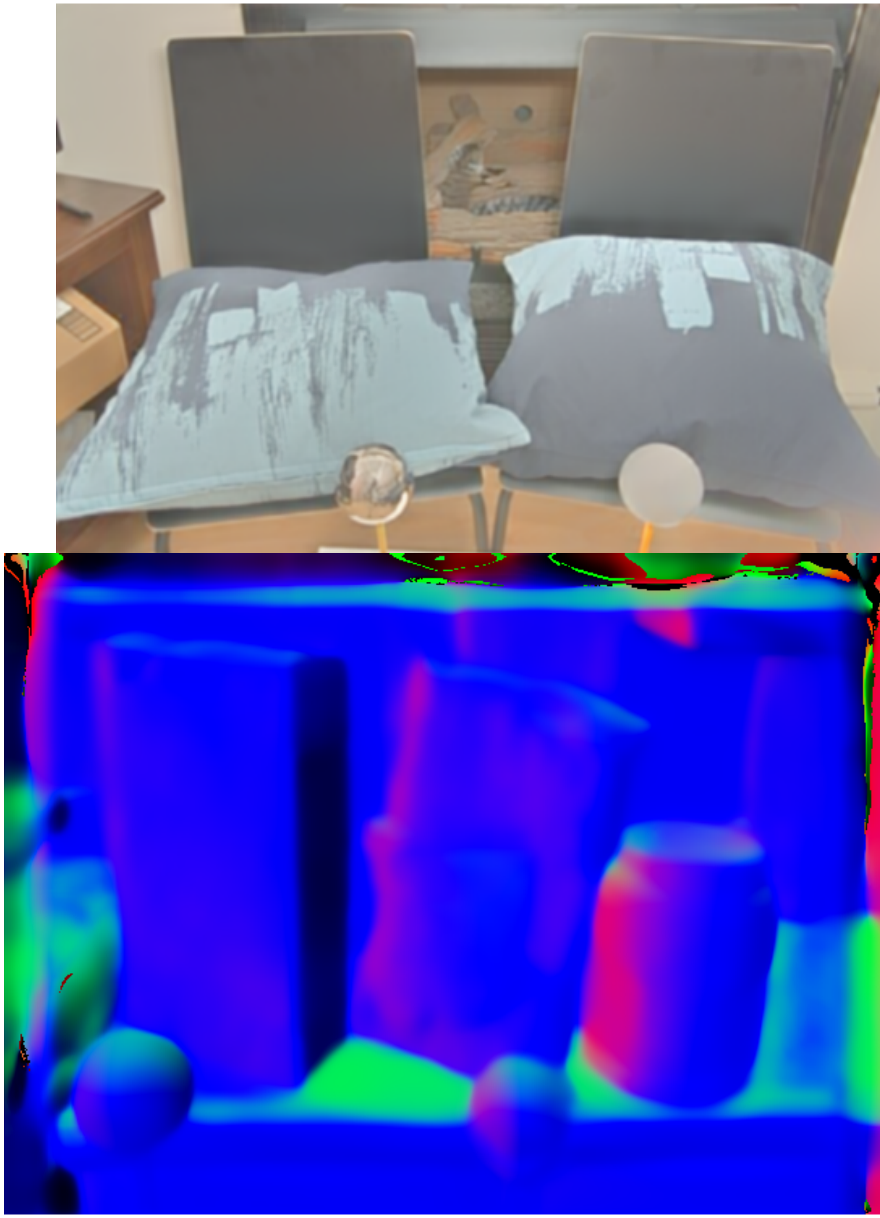
Variance control



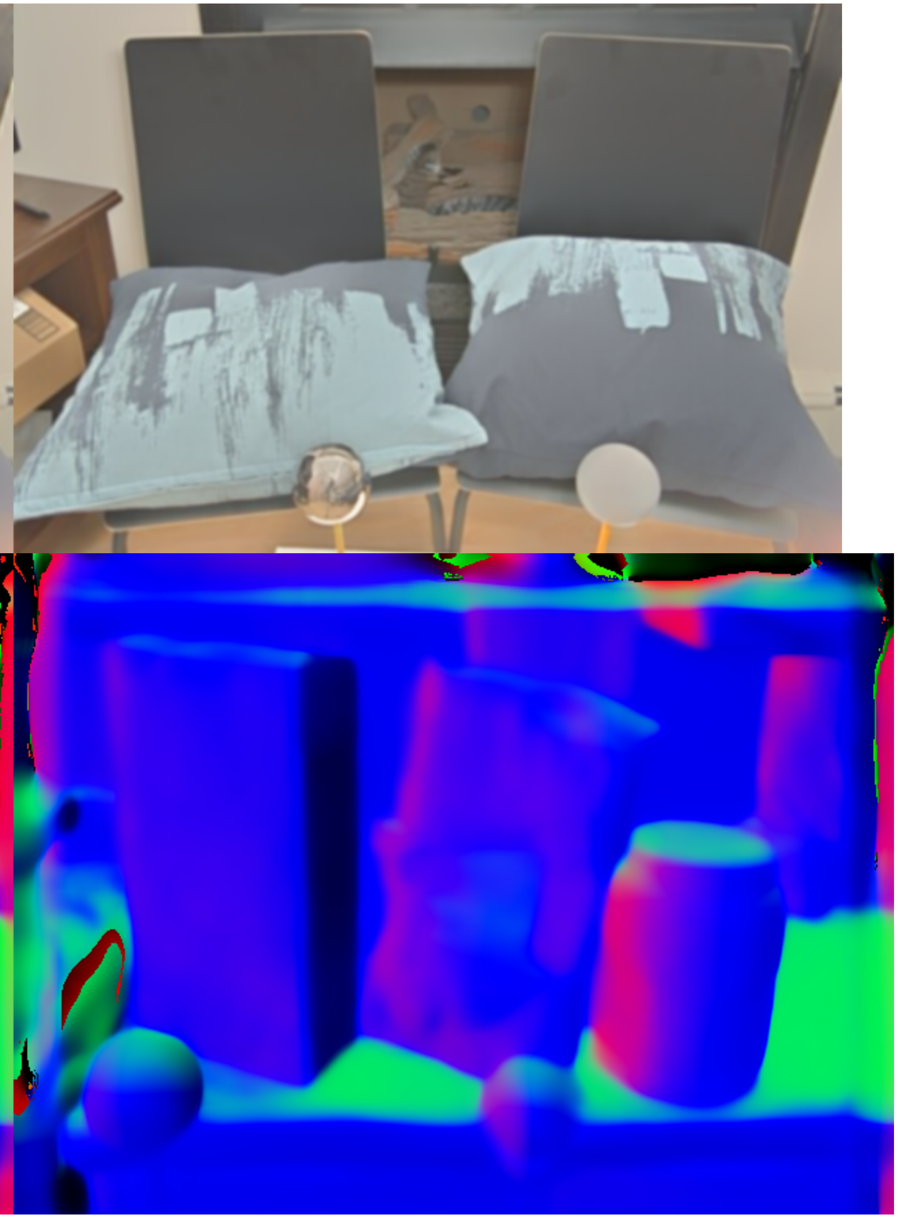
Litany of questions

- Is there an issue for depth/normal?
- How well does averaging suppress effects of illumination?
  - compare NLX with VCX (X={A, N, D})
  - 25 estimates per scene, one per illumination
  - Look at var(VCX)/var(NLX) using appropriate metrics
- Is it cheating?
  - compare means
- Are some relighting models better than others?
- Are some scenes harder than others?

Is it cheating?



Mean(NLX)



Mean(VCX)

Litany of questions

- Is there an issue for depth/normal?
- How well does averaging suppress effects of illumination?
  - compare NLX with VCX (X={A, N, D})
  - 25 estimates per scene, one per illumination
  - Look at var(VCX)/var(NLX) using appropriate metrics
- Is it cheating?
  - compare means
- Are some relighting models better than others?
- Are some scenes harder than others?

YES

VCX for the general case

- Relighting images well is very hard
  - parametrizing illumination fields is at the core
  - MIT multi-illum is special because it provides a kind of parametrization
    - for some lights

Hijacking knowledge

- StyleGAN2 is a network that
  - accepts random vectors
  - produces very convincing face images (and some others; churches, etc)
  - is trained by adversarial procedures
- This process can be “inverted”
  - GAN-inversion: given face, what random number made it?
- Pretrained models like StyleGAN2 “know” a lot
  - established literature around the idea that StyleGAN2 outputs are faces
    - pretty much whatever you do to the input

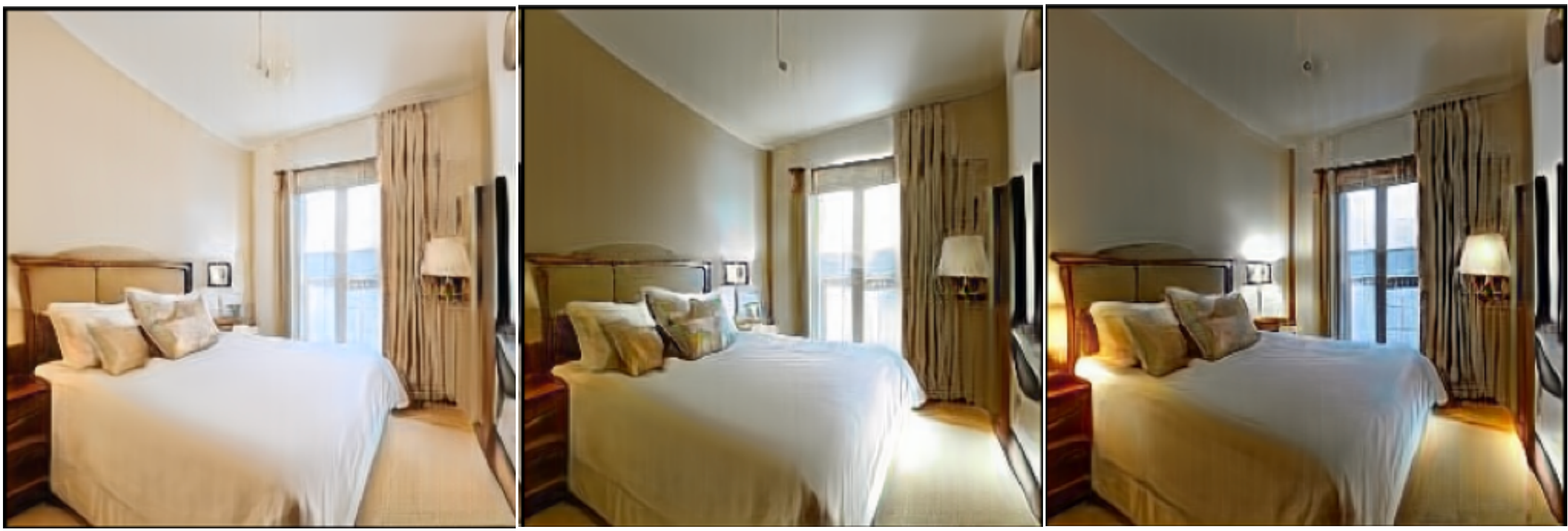
Relighting synthetic scenes

- Significant literature based on “inverse graphics”
  - Impute: geometry, albedo, luminaires; change luminaires; render
  - Zhengqin Li thesis, 2022
  - But this involves CGI,
    - which we don’t trust and
    - may not be available
- StyleGAN Judo
  - Search latent space of a generative model to find directions that
    - change image
    - don’t change computed albedo
  - for free, resurfacing
    - change image
    - don’t change shading

StyLitGAN Relighting

Luminaire aware

Luminaire aware



StyLitGAN Resurfacing

Real Images

- Problem:
  - who cares about normals/depth/albedo of generated images?
- Idea
  - apply GAN inversion to real image
  - then fiddle with lighting
  - THIS DOESN"T WORK
    - GAN inversion doesn't actually get you the image you started with

Current GAN inverters don’t

Method	LSUN Bedroom		CelebA-HQ Faces	
	MSE	LPIPS	MSE	LPIPS
ALAE	0.330	0.65	0.150	0.32
IDInvert	0.113	0.41	0.061	0.22
Psp	0.099	0.34	0.034	0.16
GHFeat (CVPR 2022)	0.068	NA	0.046	NA
PadInv (ECCV 2022)	0.054	0.21	0.021	0.10
StyleGAN2 Optim	0.17	0.42	0.020	0.009

Make it so - inversion

Make it so: Flawless Inversion

Method	LSUN Bedroom		CelebA-HQ Faces	
	MSE	LPIPS	MSE	LPIPS
ALAE	0.330	0.65	0.150	0.32
IDInvert	0.113	0.41	0.061	0.22
Psp	0.099	0.34	0.034	0.16
GHFeat (CVPR 2022)	0.068	NA	0.046	NA
PadInv (ECCV 2022)	0.054	0.21	0.021	0.10
StyleGAN2 Optim	0.17	0.42	0.020	0.009
Make it So – Simple (ours)	0.002	0.05	NA	NA
Make it So – Final (ours)	0.002	0.03	NA	NA

Image

Invert

Relights

Summary

- Big variance in (depth, normal, albedo) from lighting
- It affects strongest current methods quite severely
- Can be controlled by relighting and averaging
  - How well?
- Good relighting is important
  - and we are beginning to be quite good