RGB↔X: Image Decomposition and Synthesis Using Material- and Lighting-aware Diffusion Models

Zheng Zeng, Valentin Deschaintre, Iliyan Georgiev,
Yannick Hold-Geoffroy, Yiwei Hu, Fujun Luan, Ling-Qi Yan, Miloš Hašan

Adobe

University of California Santa Barbara

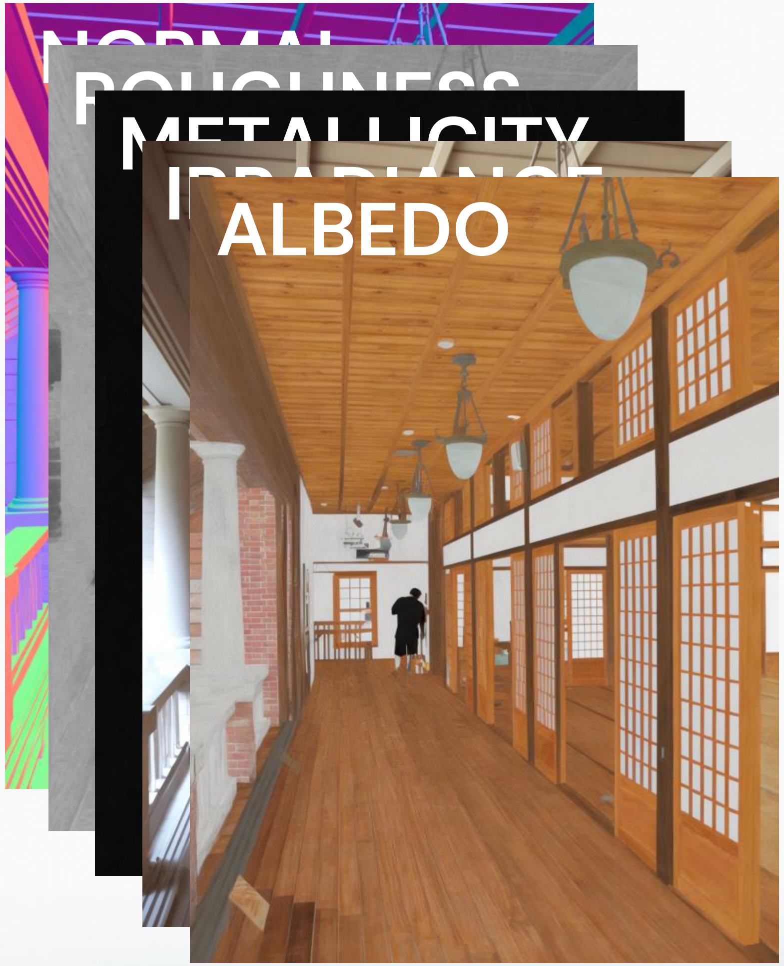
SIGGRAPH 2024
DENVER+ 28 JUL — 1 AUG

Background

Intrinsic decomposition

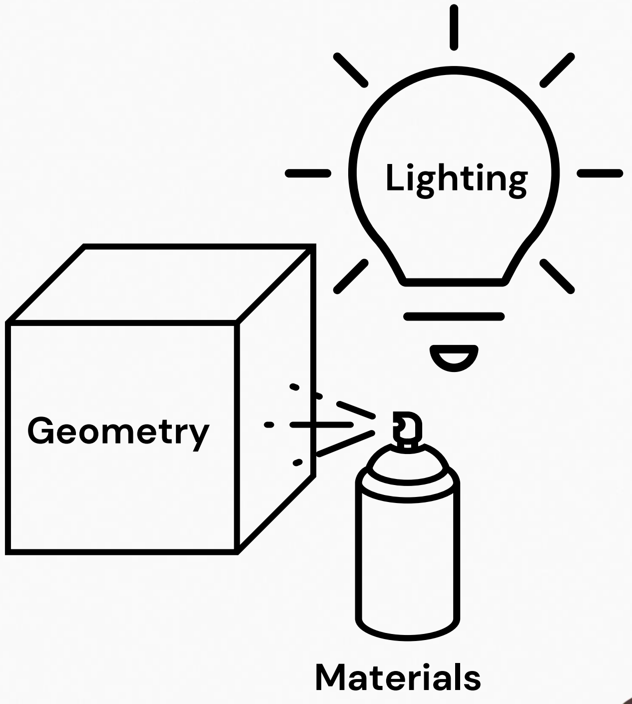
Image

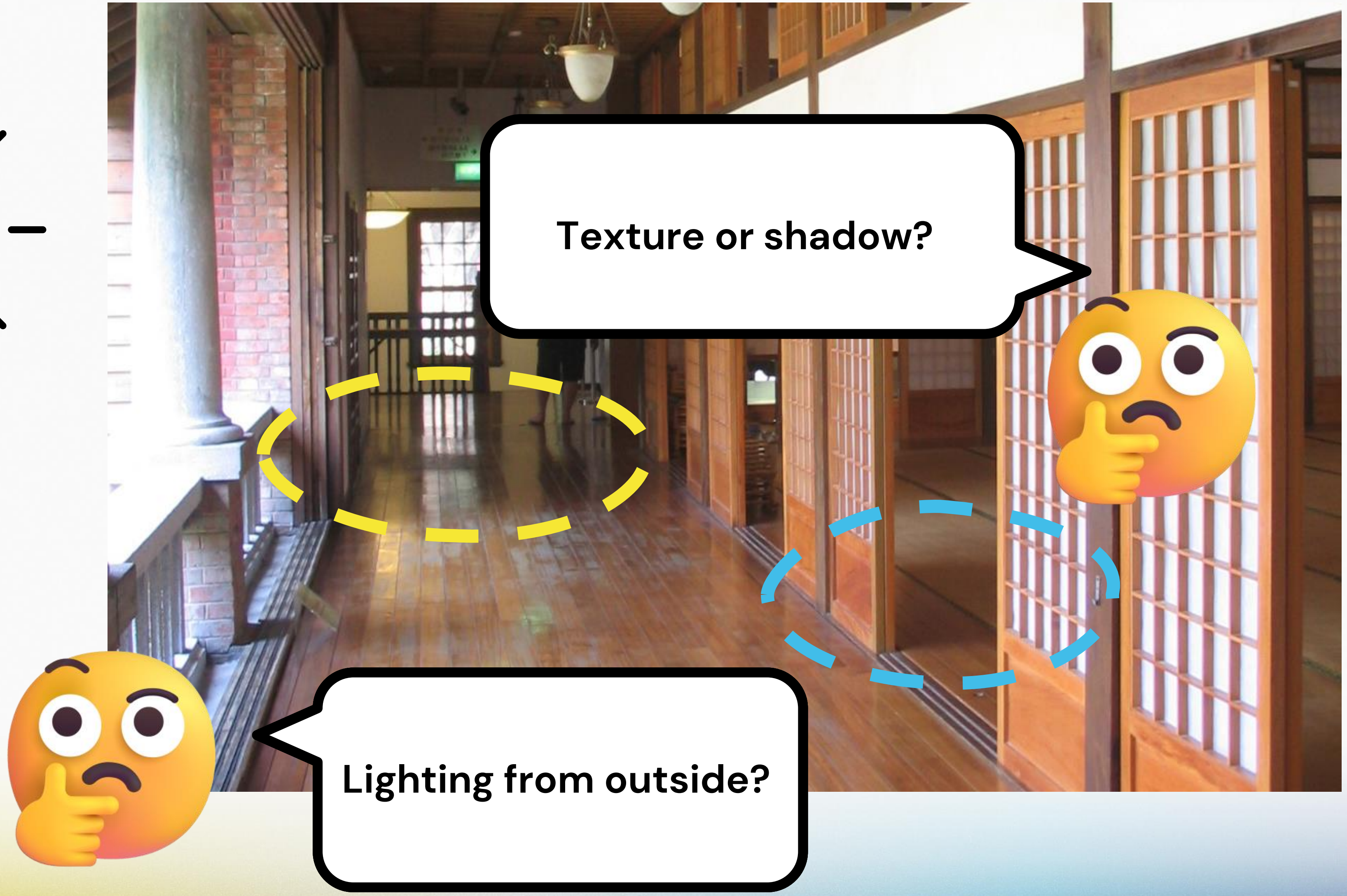
Estimate



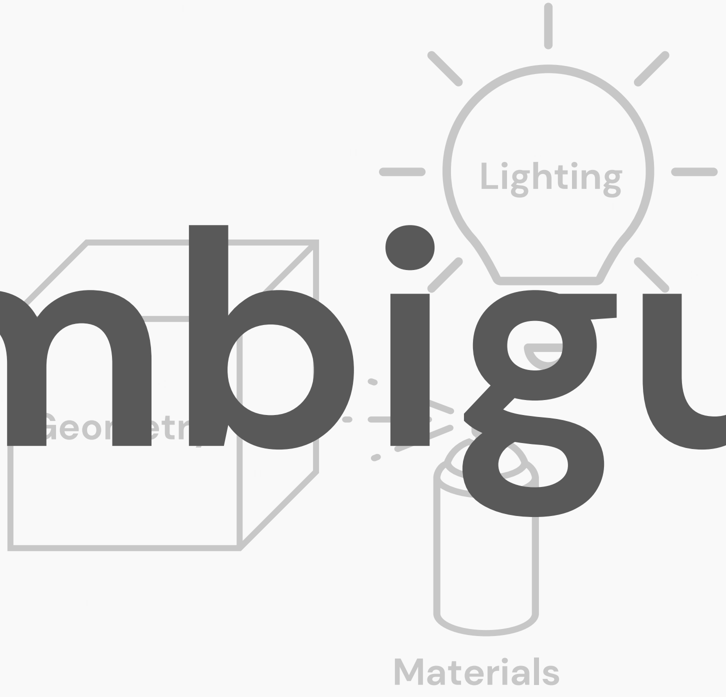
Intrinsic channels

Intrinsic decomposition is hard





Intrinsic decomposition is hard



Ambiguity

Previous work on intrinsic decomposition (albedo channel)

SIGGRAPH 2024
DENVER+ 28 JUL — 1 AUG

Image

[Zhu et al. 2022b]

Previous work on intrinsic decomposition (albedo channel)

Image

[Zhu et al. 2022b]

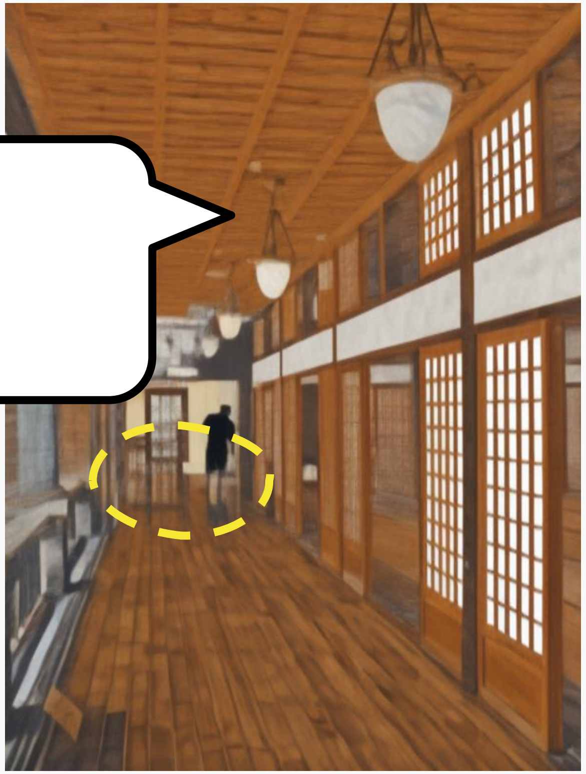
[Kocsis et al. 2023]

Previous work on intrinsic decomposition (albedo channel)

Image

[Zhu et al. 2022b]

Finetune a diffusion model

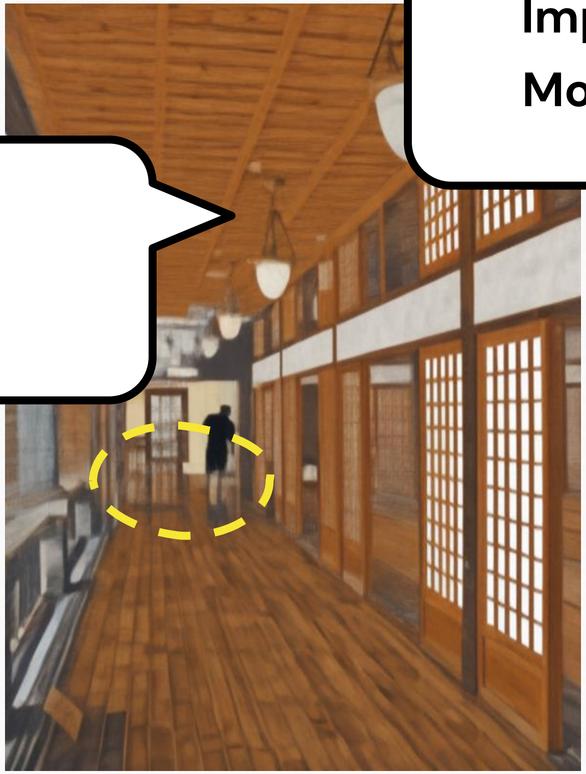


[Kocsis et al. 2023]

Previous work on intrinsic decomposition (albedo channel)

Finetune a diffusion model

[Zhu et al. 2022b]



[Kocsis et al. 2023]

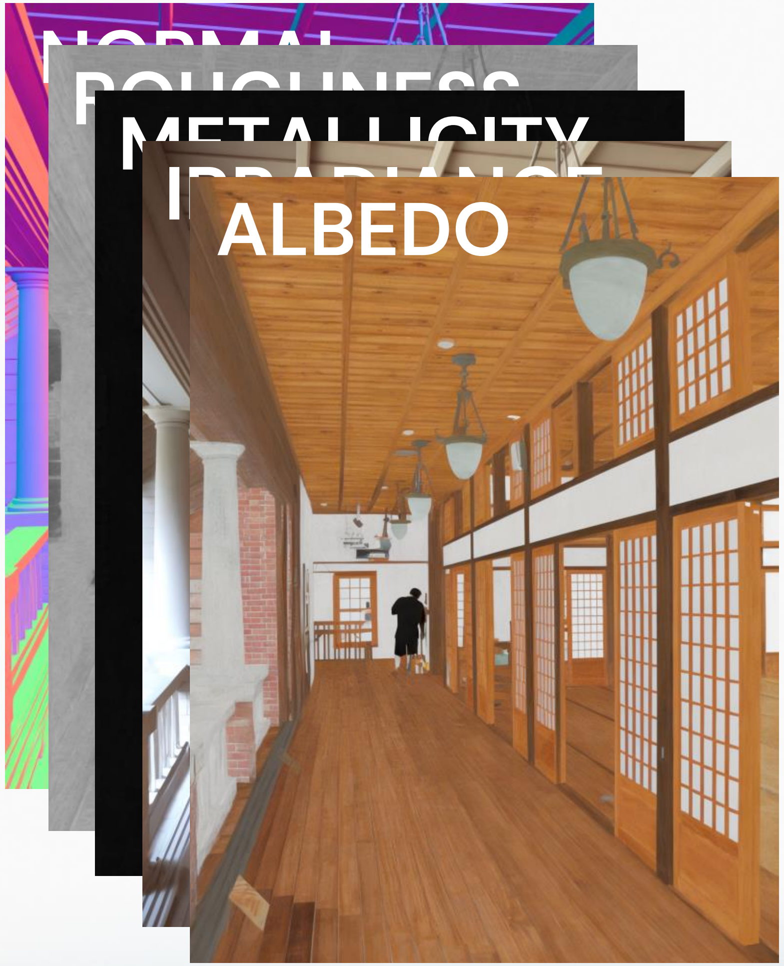
Improvements?
More channels?

We follow this direction further

Intrinsic decomposition

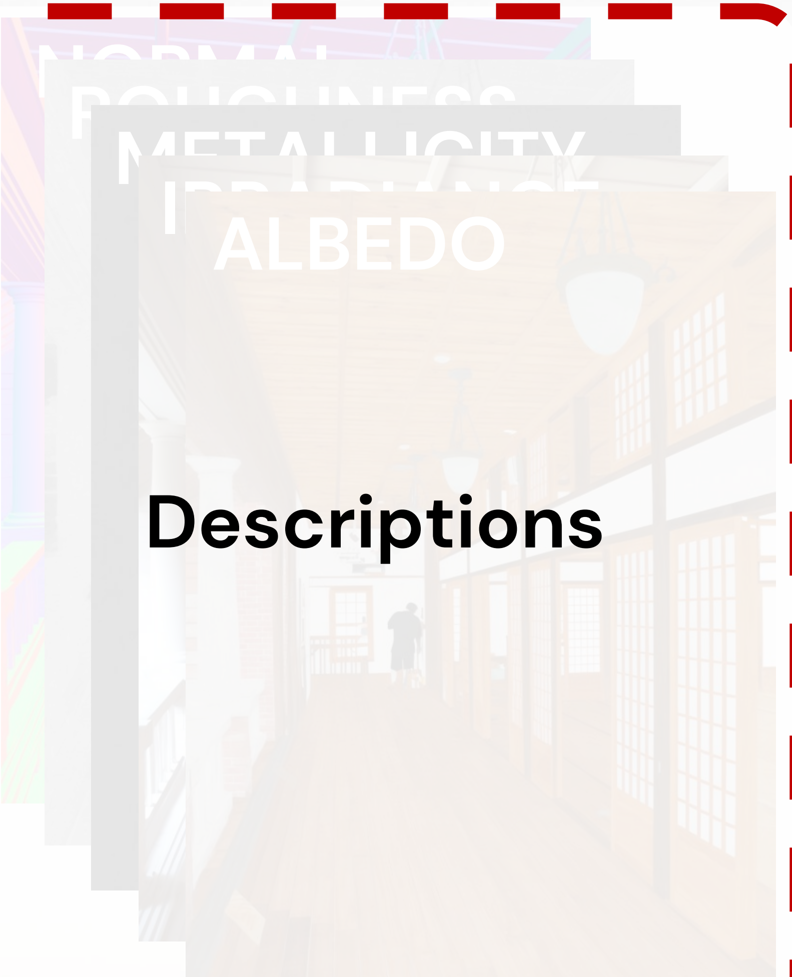
Image

Estimate



Intrinsic channels

Image synthesis



Descriptions

Intrinsic channels

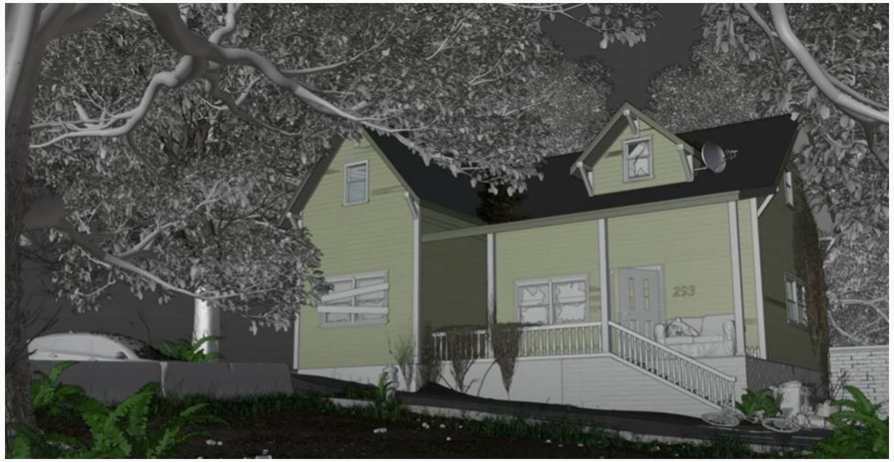
Synthesized image

Image synthesis

Physically based rendering

*Images from Andrew Price's Blender tutorial

Image synthesis



Scene description

Physically based rendering

Realistic image

- Precise and consistent
- Perfect controllability
- Requires full scene description

*Images from Andrew Price's Blender tutorial

Image synthesis

Image diffusion models

Image synthesis

"A modern kitchen"

Prompt

Image diffusion models

- Simple to use
- Hard to precisely control

Image generated by SD v3

Image synthesis

A middle ground?

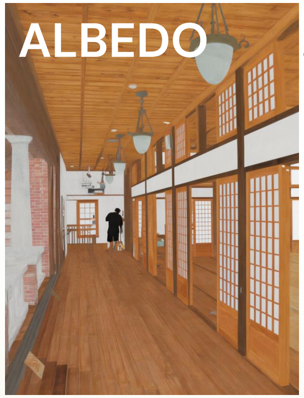
Physically based rendering

- Precise and consistent
- Easy authoring
- Requires full scene description

Image diffusion models

- Simple to use
- Hard to precisely control

Properties



Intrinsic channels

Synthesized image

Summarize the motivation

Summarize the motivation

RGB → X

Image

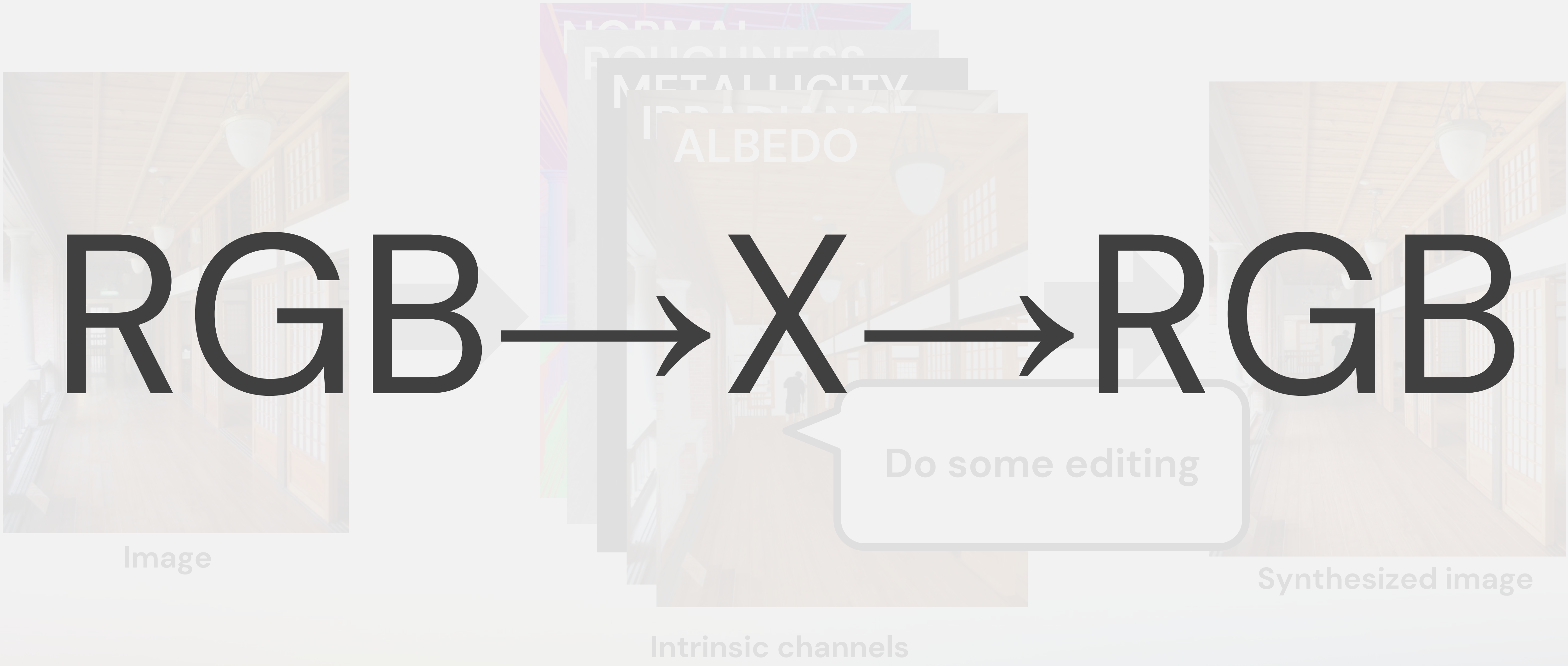
Intrinsic channels

Summarize the motivation

X→RGB

Summarize the motivation

$$\mathrm{RGB} \longrightarrow \mathrm{X} \longrightarrow \mathrm{RGB}$$

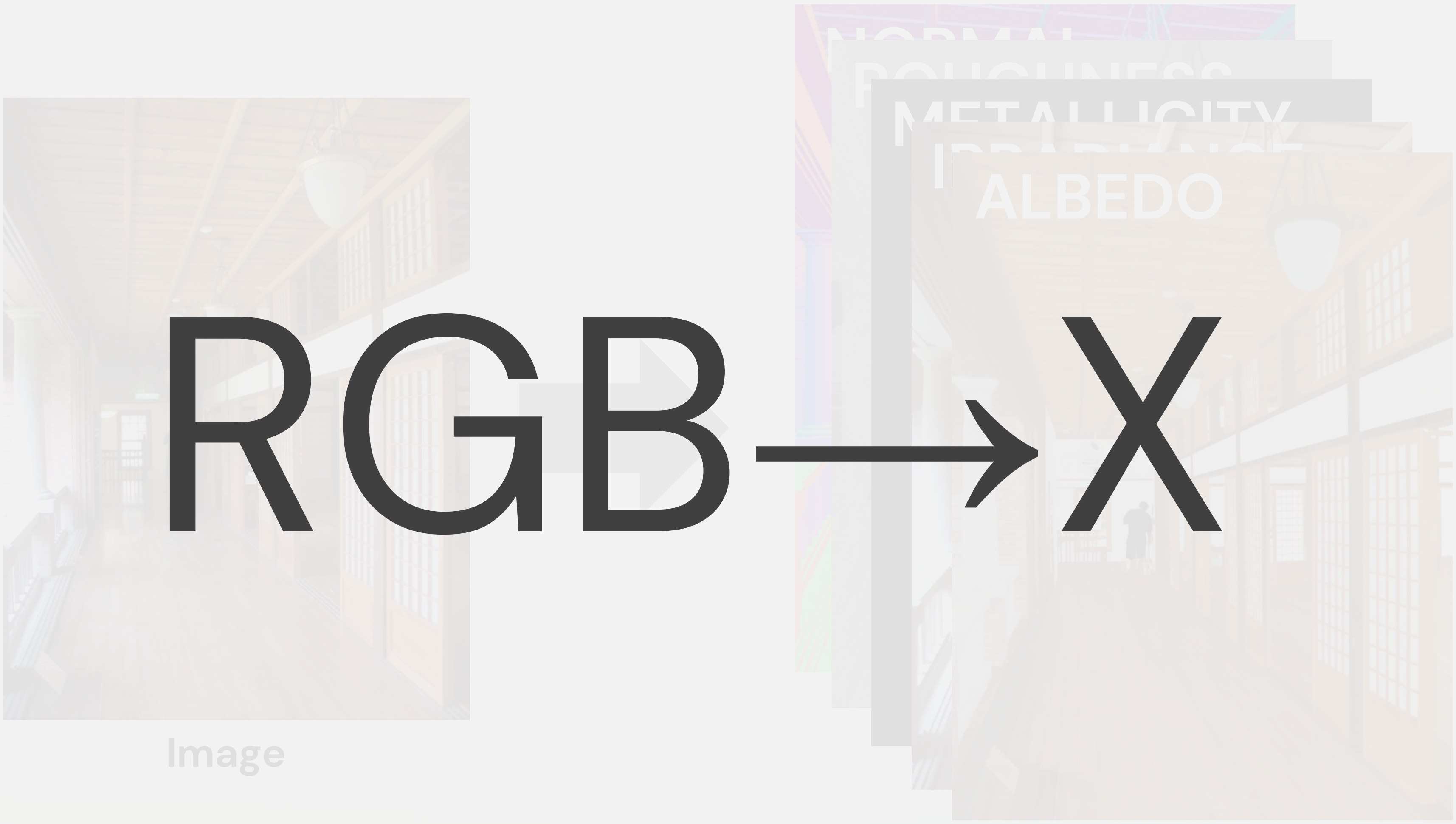


SIGGRAPH 2024

DENVER+ 28 JUL — 1 AUG

Method

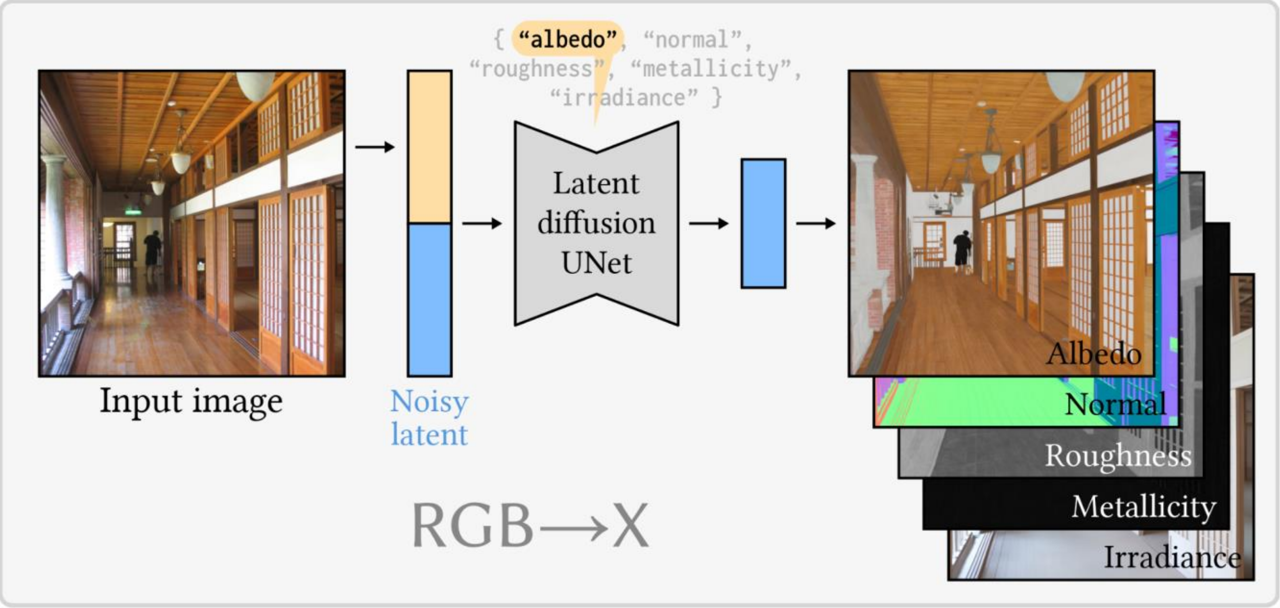
RGB→X



Image

Intrinsic channels

RGB→X: how it works?



Finetune Stable Diffusion on synthetic data

- Conditioned on image RGB
- Produce intrinsic channels X

Re-purpose "prompt" as a "switch"

- Example: given "albedo", it produces albedo
- Benefits:
  - Avoid finetuning multiple outputs – it's harder
  - Enable usage of datasets with different available channels
- More details in paper

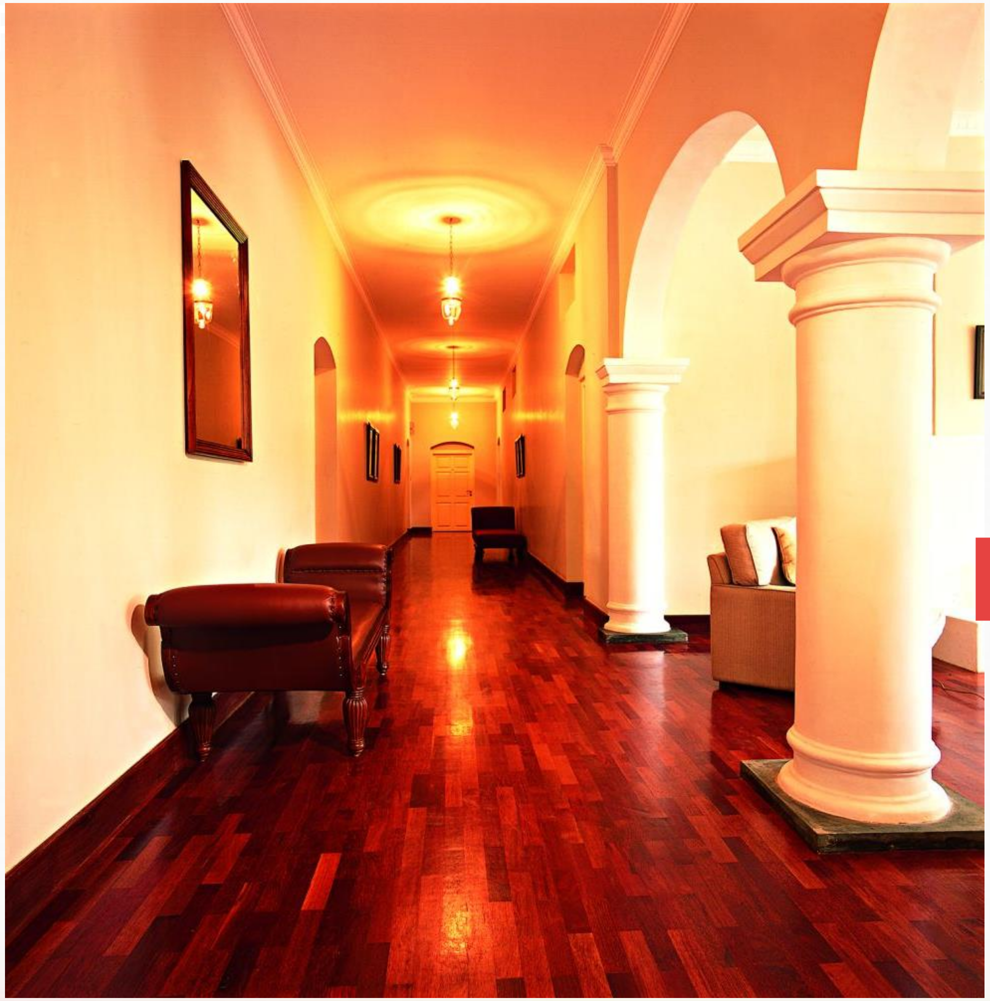
Heterogeneous synthetic training data

✓: available. ✓: available but not reliable. ✗: not available.

Dataset	Size	Albedo	Normal	Roughness	Metallic.	Irrad.
InteriorVerse	50,097	✓	✓	✓	✓	✗
Hypersim	73,819	✓	✓	✗	✗	✓
Evermotion	17,000	✓	✓	✓	✓	✗
ImageDecomp	50,000	✓	✓	✓	✓	✓

InteriorVerse Hypersim Evermotion ImageDecomp

RGB→X results

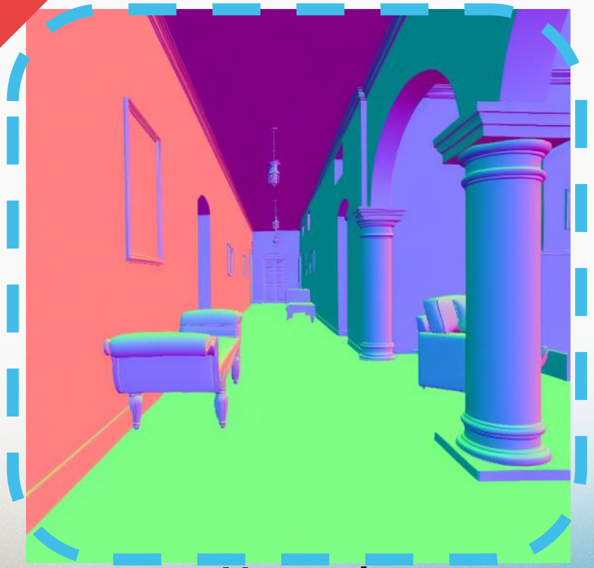


Input (real photo)

RGB→X

Albedo

Irradiance

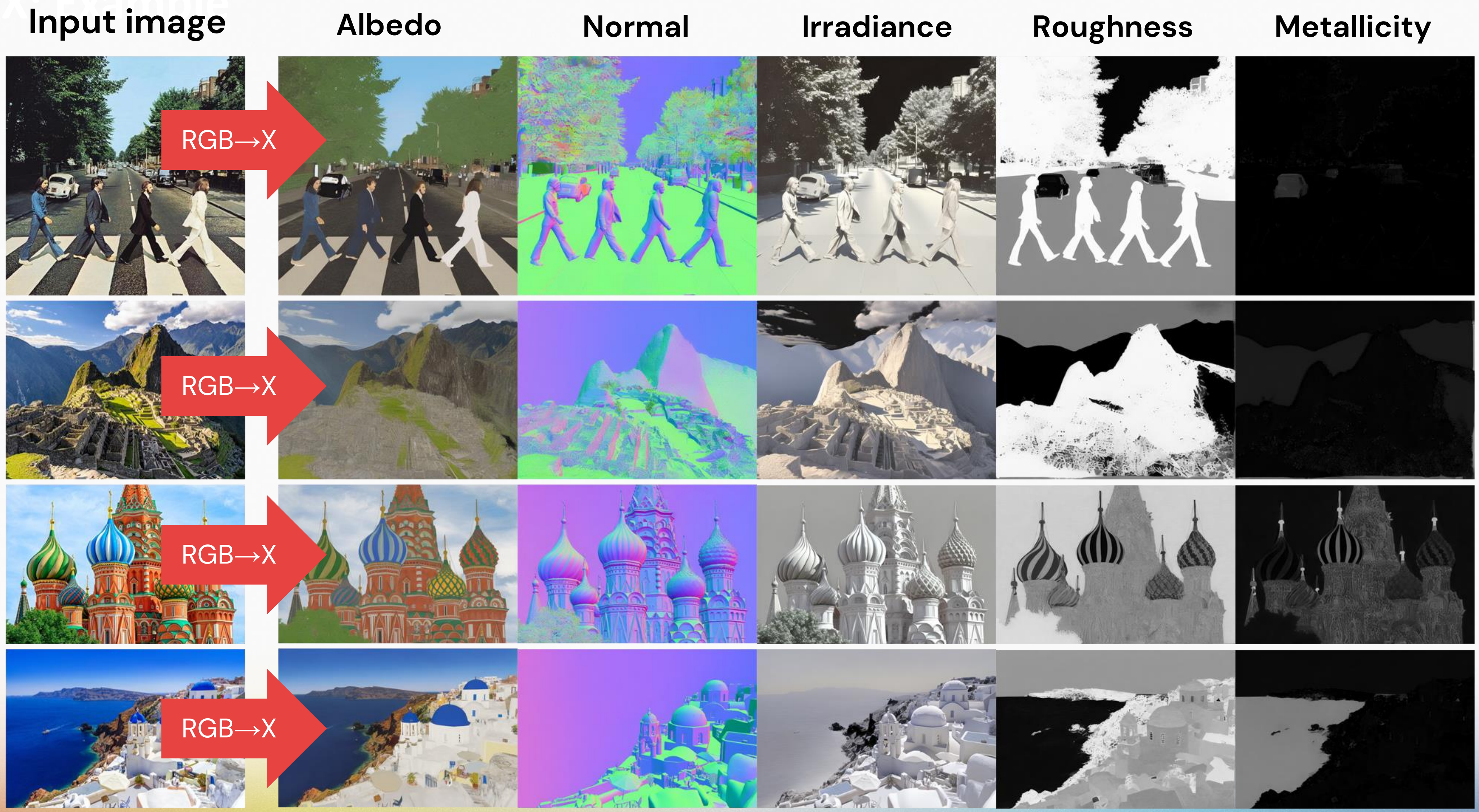


Normal

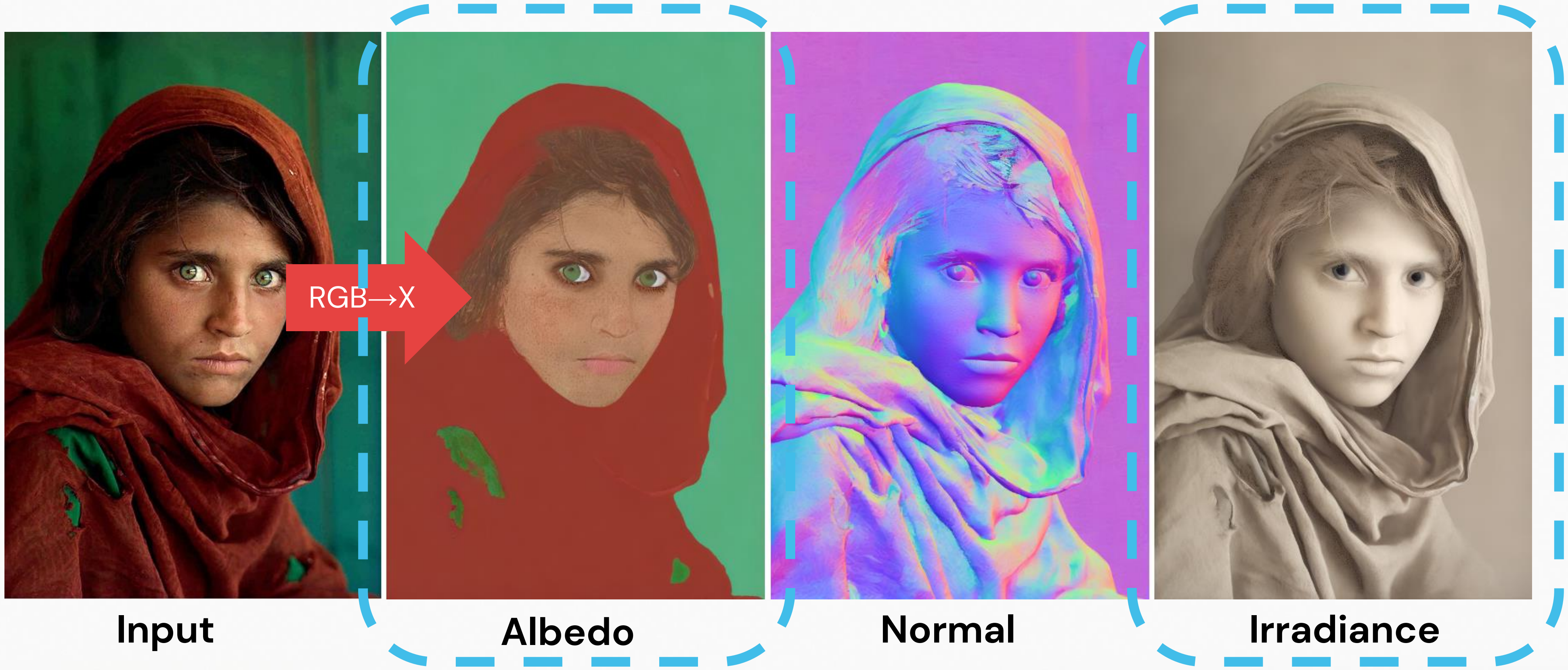
Roughness

RGB→X results (works, despite no outdoor training data)

SIGGRAPH 2024
DENVER+ 28 JUL — 1 AUG



RGB→X results (works, despite no human face training data)

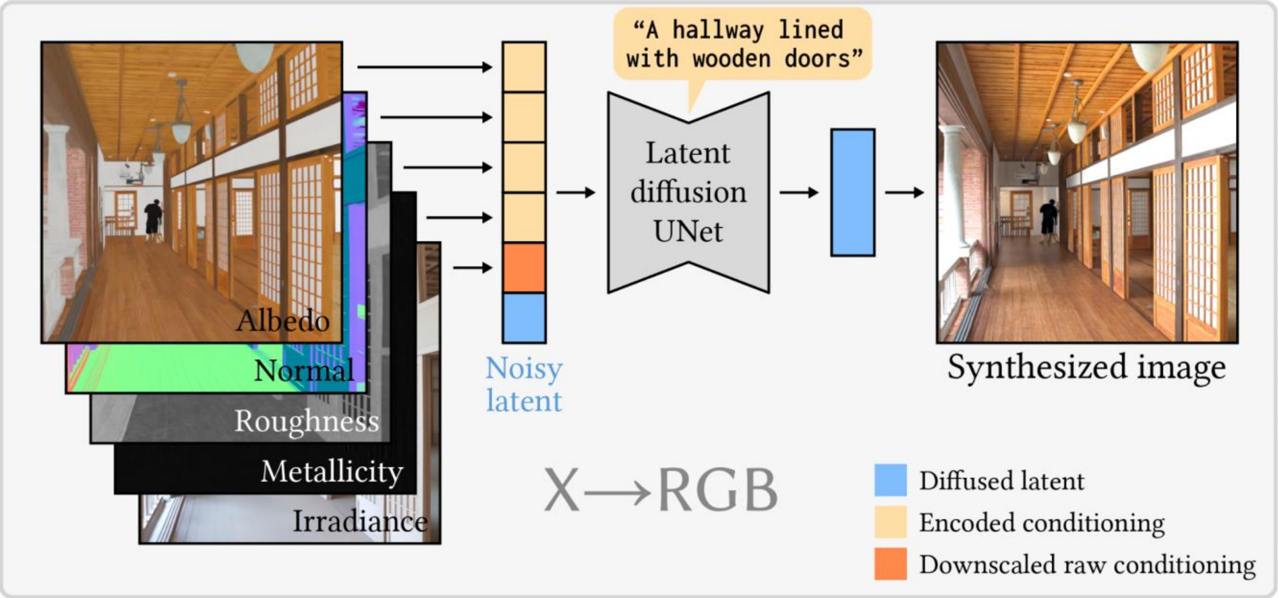


X→RGB

Intrinsic channels

Synthesized image

X→RGB: how it works



Finetune Stable Diffusion on synthetic data

- Conditioned on intrinsic channels X
- Produce image RGB

Intrinsic channel dropout strategy

- Randomly drop condition channels during training
- Benefits: this lets us
  - handle heterogeneous datasets during training
  - choose which inputs to provide at inference

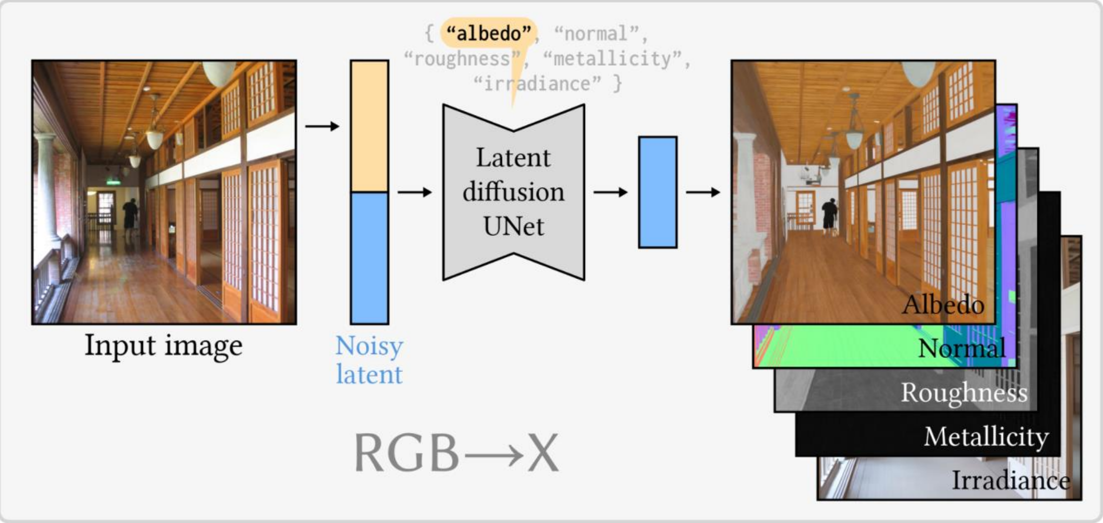
X→RGB results (comparison to classical rendering)

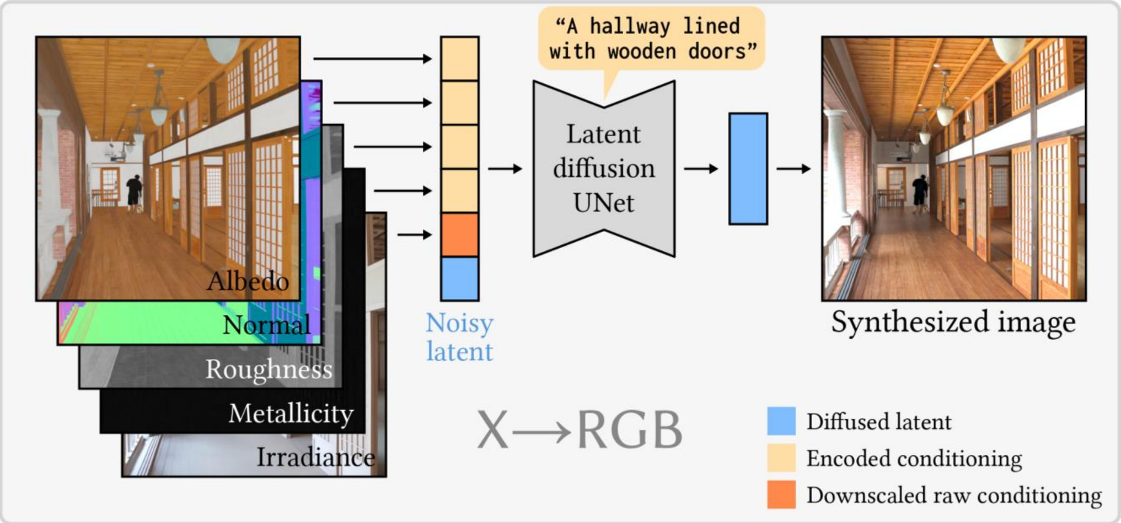
Our X→RGB result
(from intrinsic channels)

Reference classical rendering
(needs full scene)

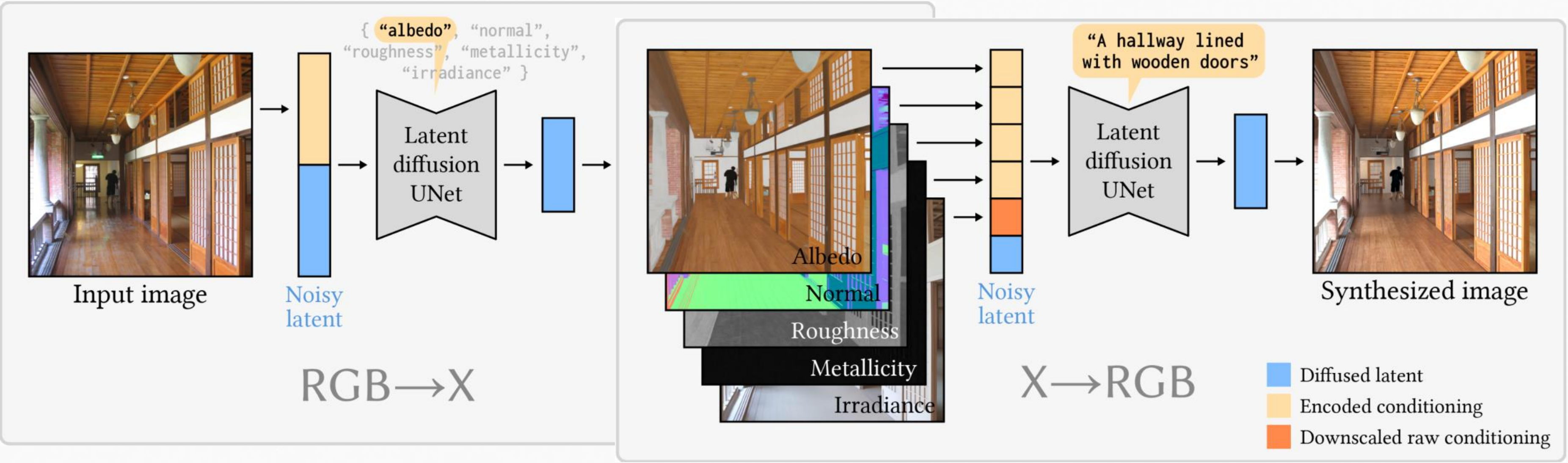
X→RGB results (material / lighting control by text prompts)

Having RGB→X and X→RGB?

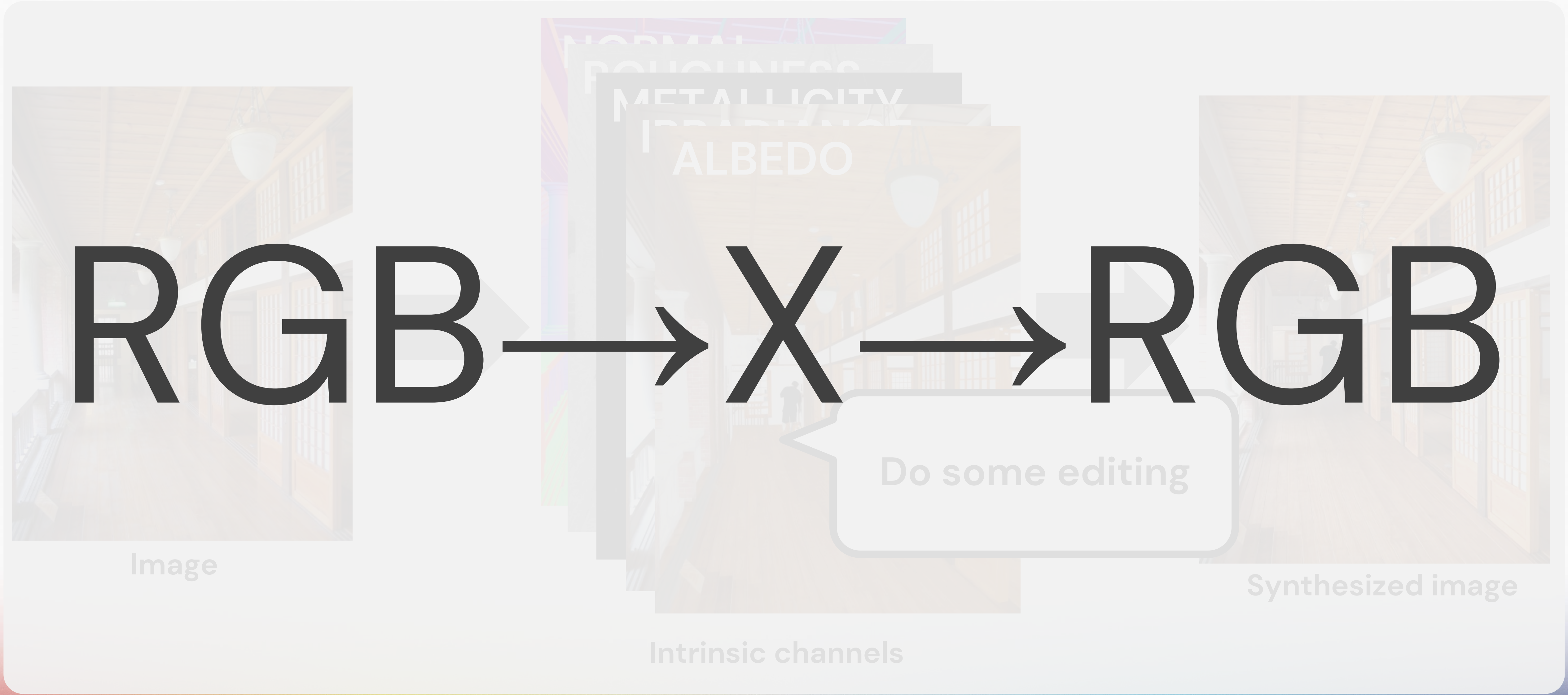




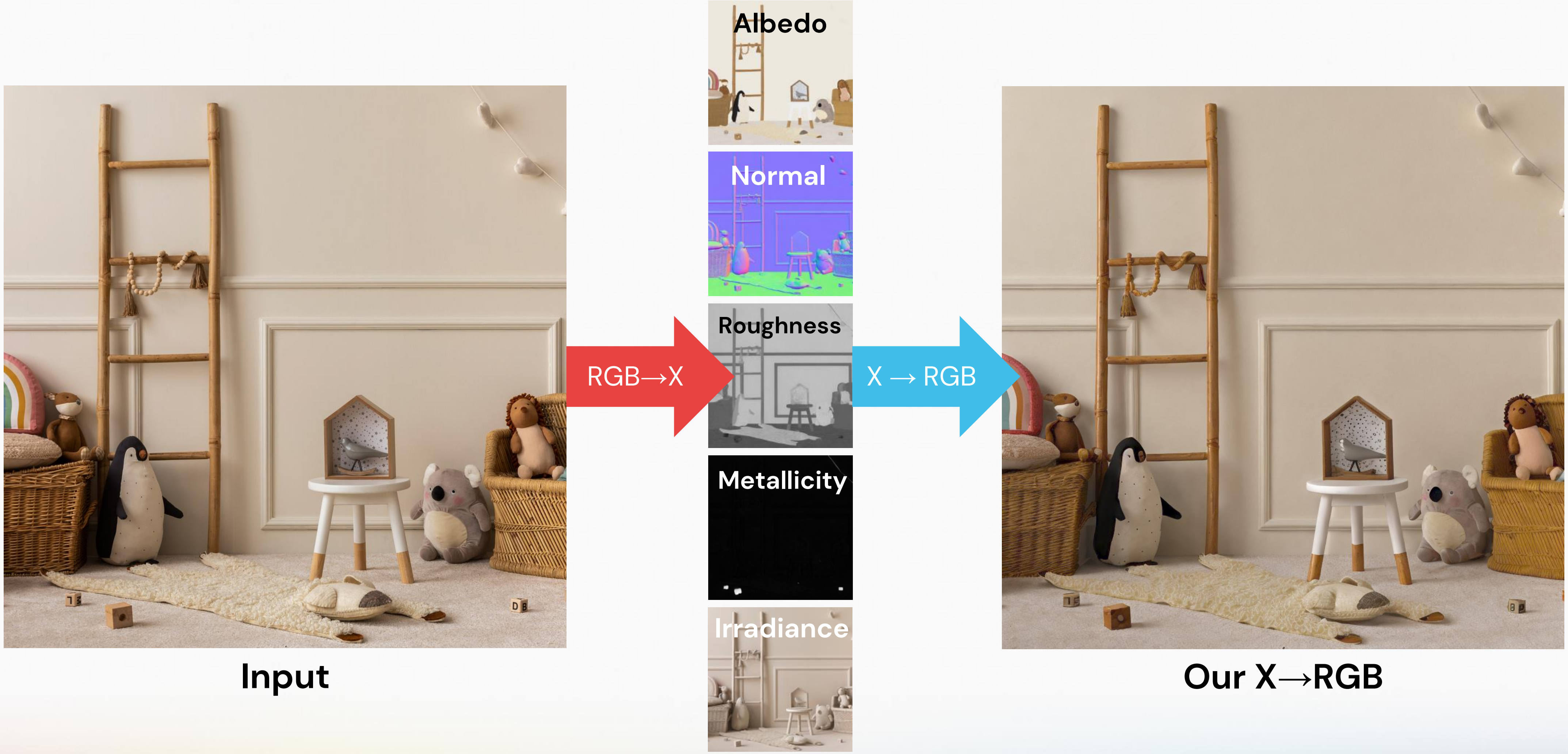
RGB→X→RGB


{ "albedo", "normal", "roughness", "metallicity", "irradiance" }
Latent diffusion UNet
Input image
Noisy latent
RGB⟶X
Albedo
Normal
Roughness
Metallicity
Irradiance
"A hallway lined with wooden doors"
Latent diffusion UNet
Noisy latent
Synthesized image
X⟶RGB
Diffused latent
Encoded conditioning
Downscaled raw conditioning

RGB → X → RGB


METALLICITY
ALBEDO
Do some editing
Image
Synthesized image
Intrinsic channels

RGB→X→RGB results



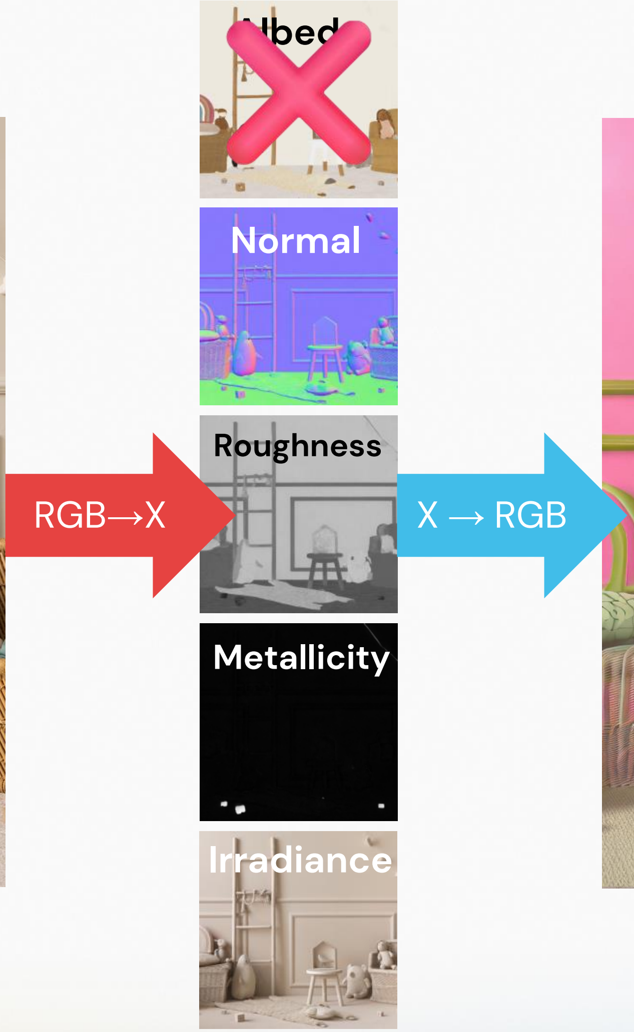
Input

Our X→RGB

Our intrinsic channels X

RGB→X→RGB results

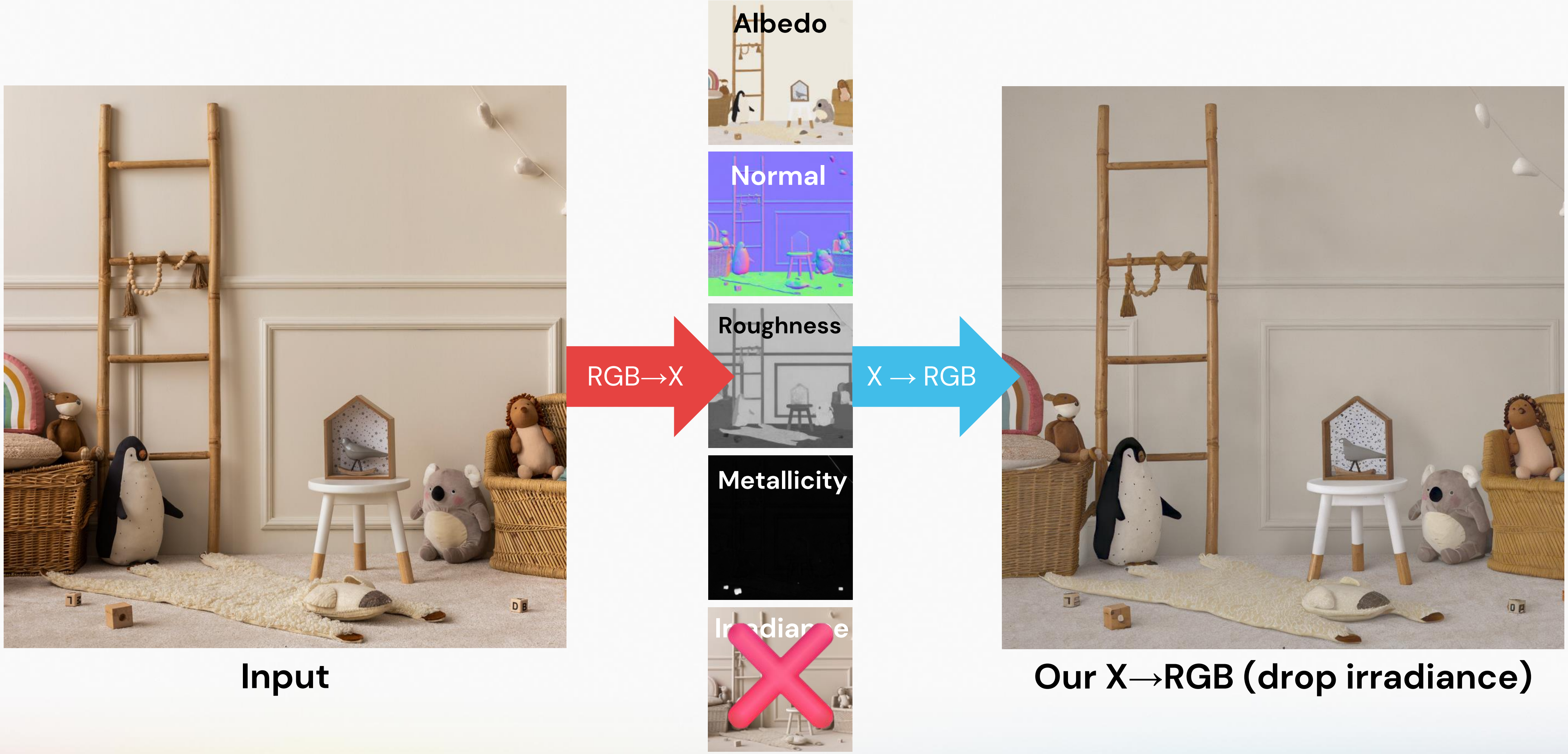
Input



Our X→RGB (drop albedo)

Our intrinsic channels X

RGB→X→RGB results

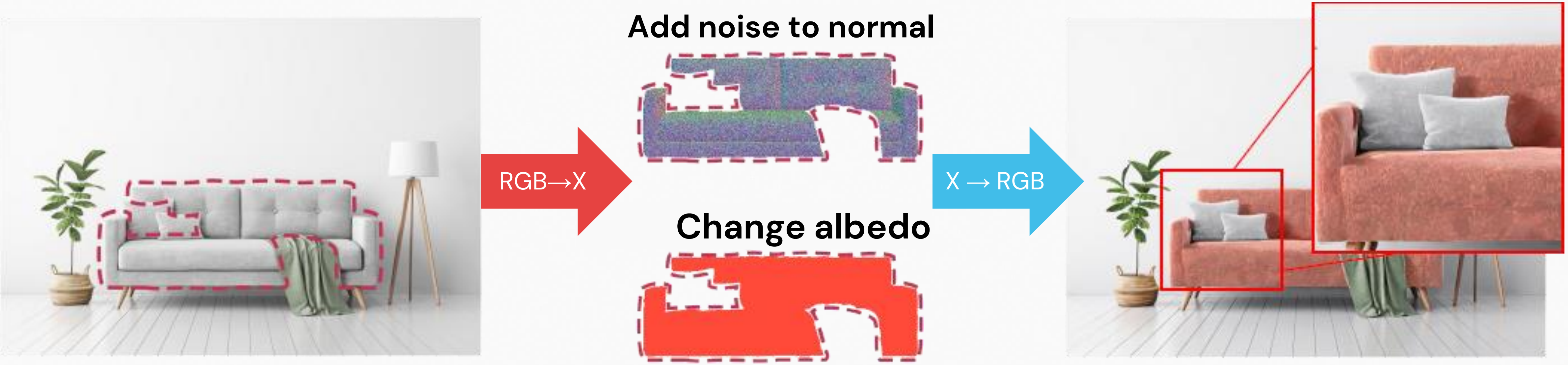


Input

Our intrinsic channels X

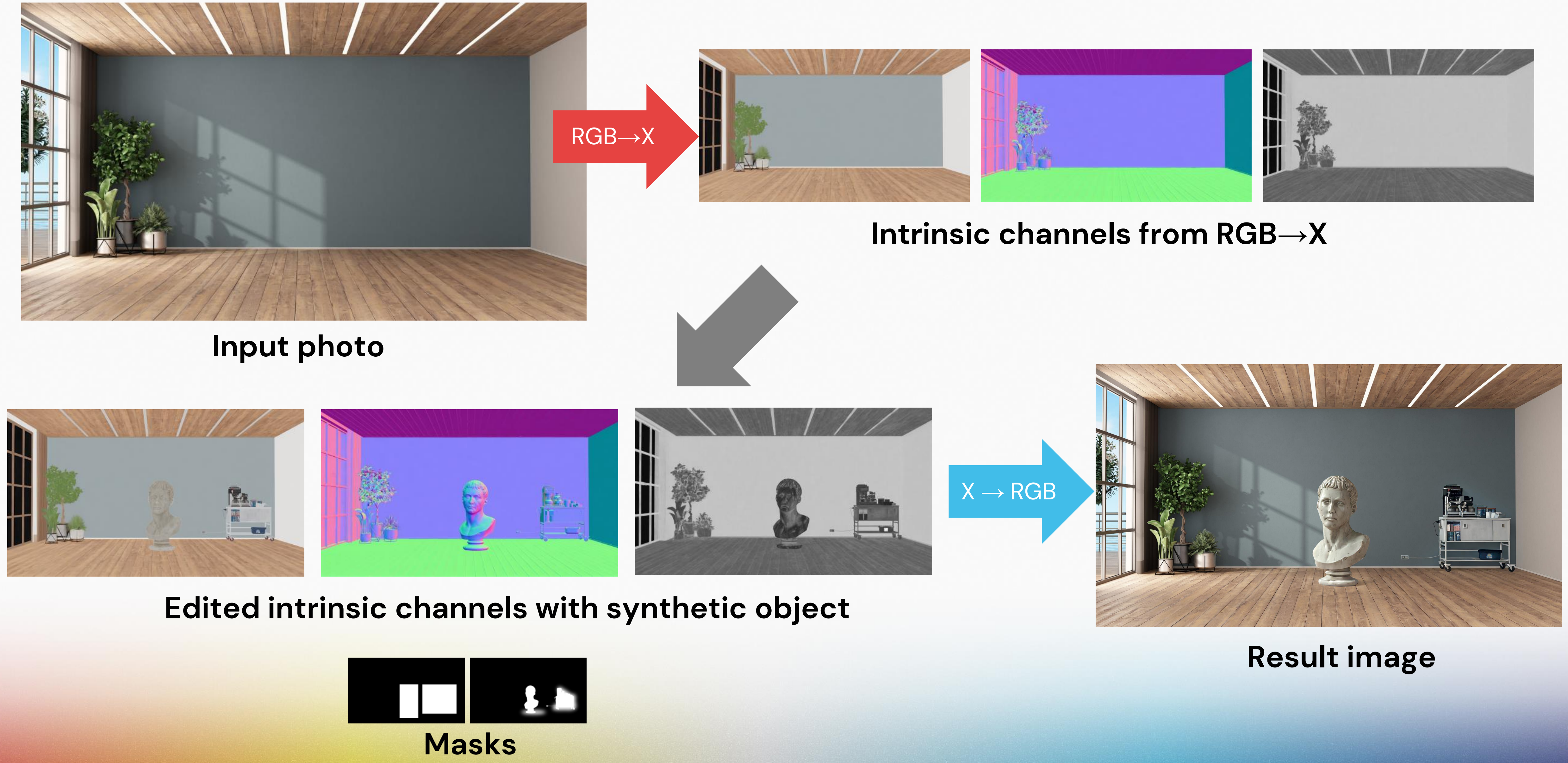
Our X→RGB (drop irradiance)

RGB→X→RGB application: material editing

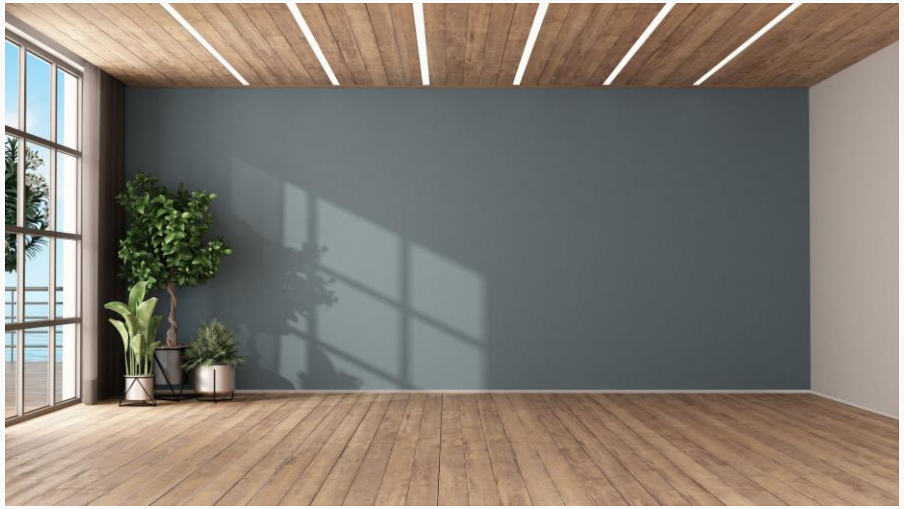


Input image with mask

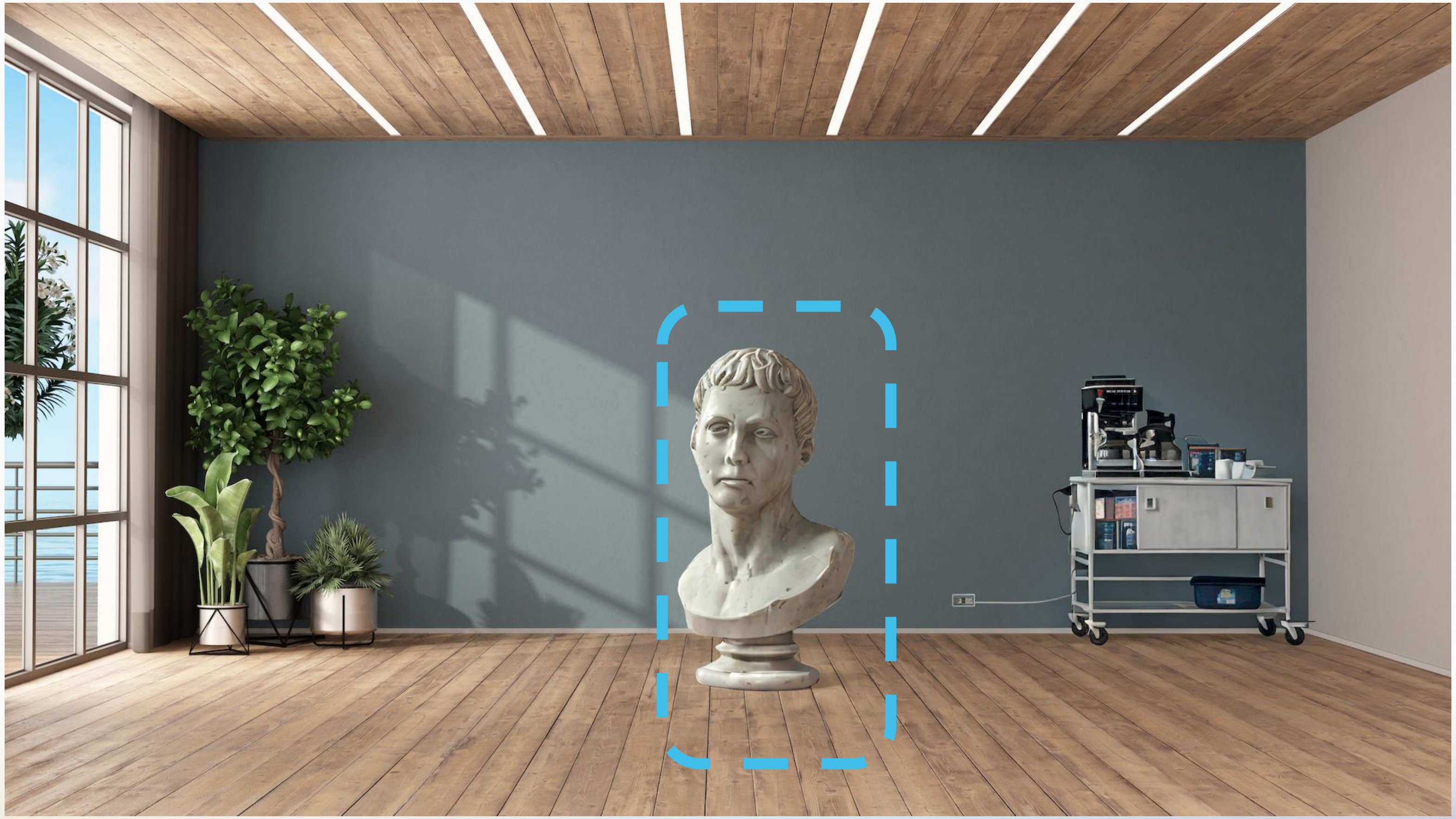
RGB→X→RGB application: synthetic object insertion



RGB→X→RGB application: synthetic object insertion



Input photo

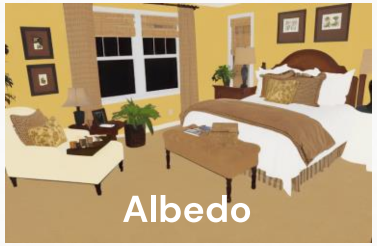


Result image

RGB→X→RGB application: relighting

Input photo

RGB→X



X → RGB

Relit

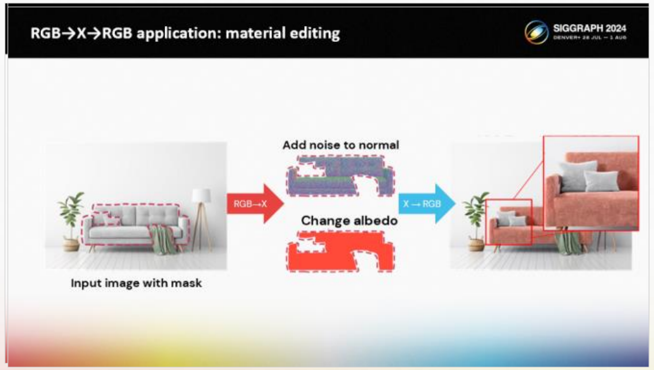
Thanks to Julien Philip for contributing on this result.

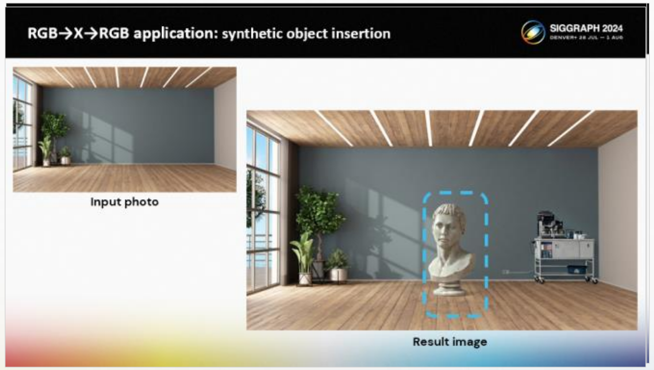
SIGGRAPH 2024
DENVER+ 28 JUL — 1 AUG

Summary

Summary

- A unified diffusion framework for
  - intrinsic channel estimation from images (termed RGB→X) and
  - synthesizing realistic images from such channels (X→RGB)
- RGB→X→RGB enables
  - Material editing, object insertion, relighting





Project page here

RGB↔X: Image Decomposition and Synthesis Using Material- and Lighting-aware Diffusion Models

Zheng Zeng, Valentin Deschaintre, Iliyan Georgiev,
Yannick Hold-Geoffroy, Yiwei Hu, Fujun Luan, Ling-Qi Yan, Miloš Hašan

Thank you!