30 JULY – 3 AUGUST Los Angeles
SIGGRAPH 2017

Modeling Surface Appearance from a Single Photograph using Self-Augmented Convolutional Neural Networks

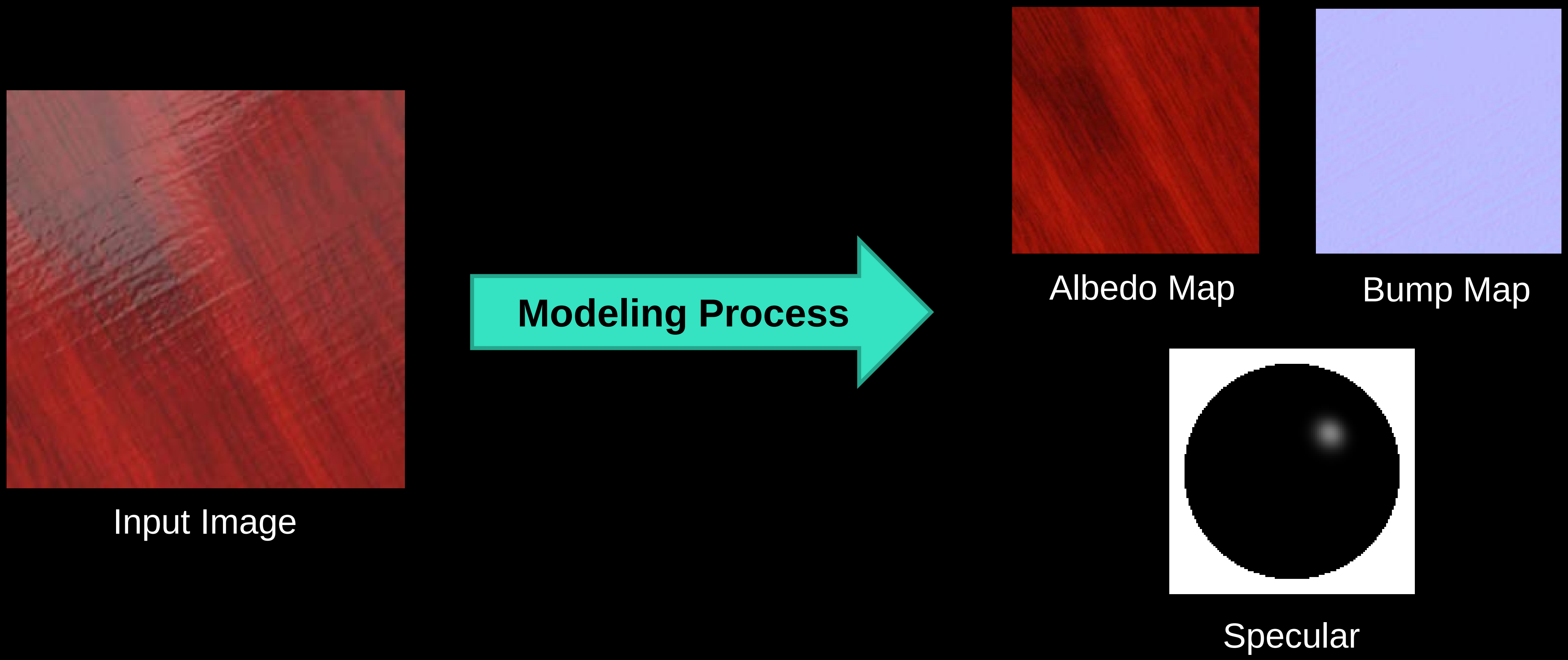
Xiao Li[1,2] Yue Dong[2] Pieter Peers[3] Xin Tong[2]

[1] University of Science and Technology of China

[2] Microsoft Research, Beijing

[3] College of William & Mary

Appearance Modeling from Single Image



Artists’ Solution

Input Image

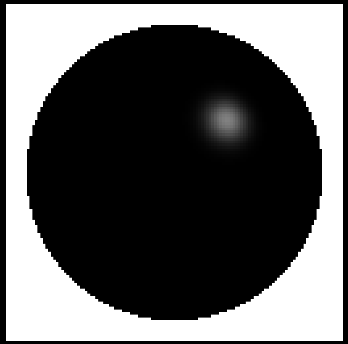
Professional software

Manual Interactions

Experienced artist

Albedo Map

Bump Map



Specular

Our Goal

Input Image

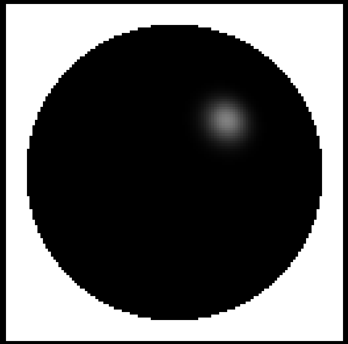
End to End System

Automatic Process

Non-expert Users

Albedo Map

Bump Map

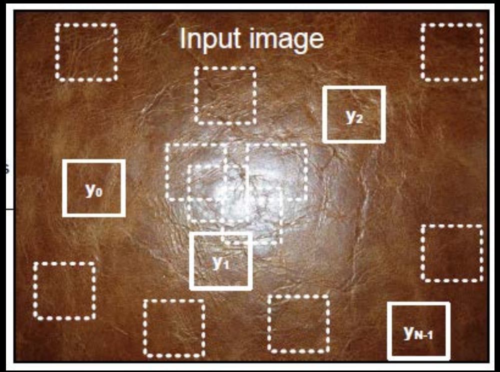


Specular

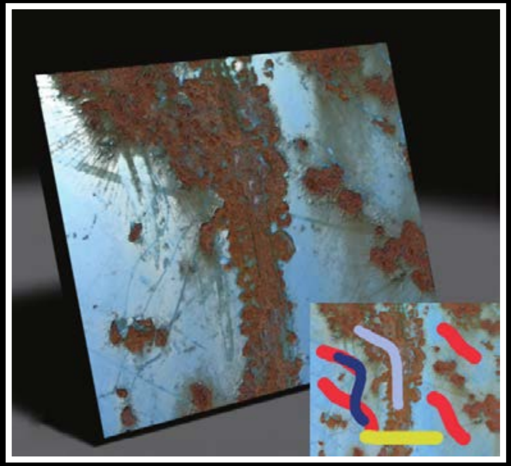
Related Work

Single image appearance modeling

- Active illumination / Known lighting
  - [Wang 2016]; [Xu 2016]
- Stationary / Stochastic Textures
  - [Wang 2011]; [Aittala 2016]
- Diffuse / homogeneous BRDF
  - [Barron 2015]; [Shi 2017]
- Manual interaction
  - [Dong 2011]



[Aittala 2016]



[Dong 2011]

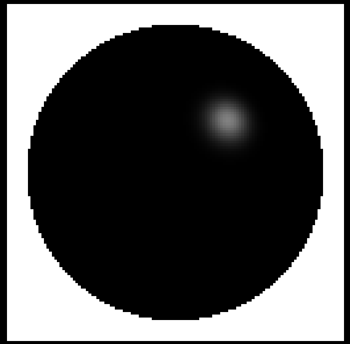
Modeling Appearance by CNN

Input Image

Convolutional Neural Network

Albedo Map

Bump Map



Homogeneous Specular
(Ward Model)

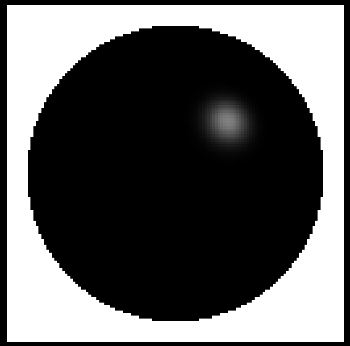
Obtaining Labeled Data is HARD!

Input Image

Labeled data pair

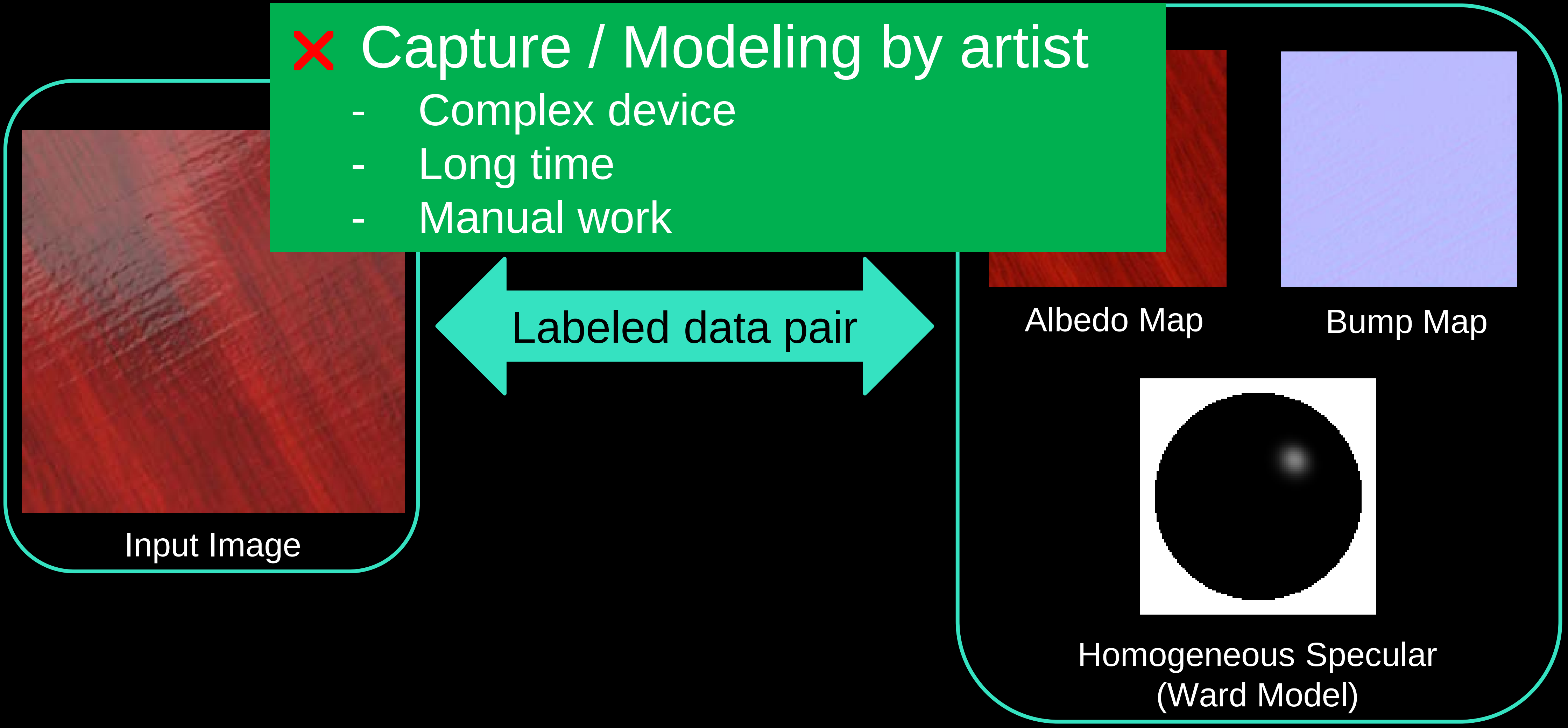
Albedo Map

Bump Map



Homogeneous Specular
(Ward Model)

Obtaining Labeled Data is HARD!



Obtaining Labeled Data is HARD!

❌ Capture / Modeling by artist
- Complex device
- Long time
- Manual work

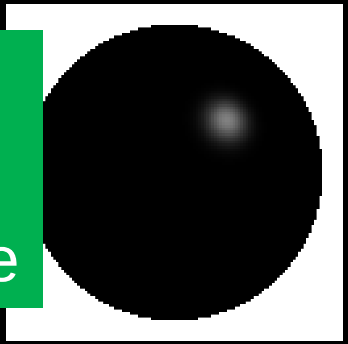
Input Imag

Labeled data pair

Albedo Map

Bump Map

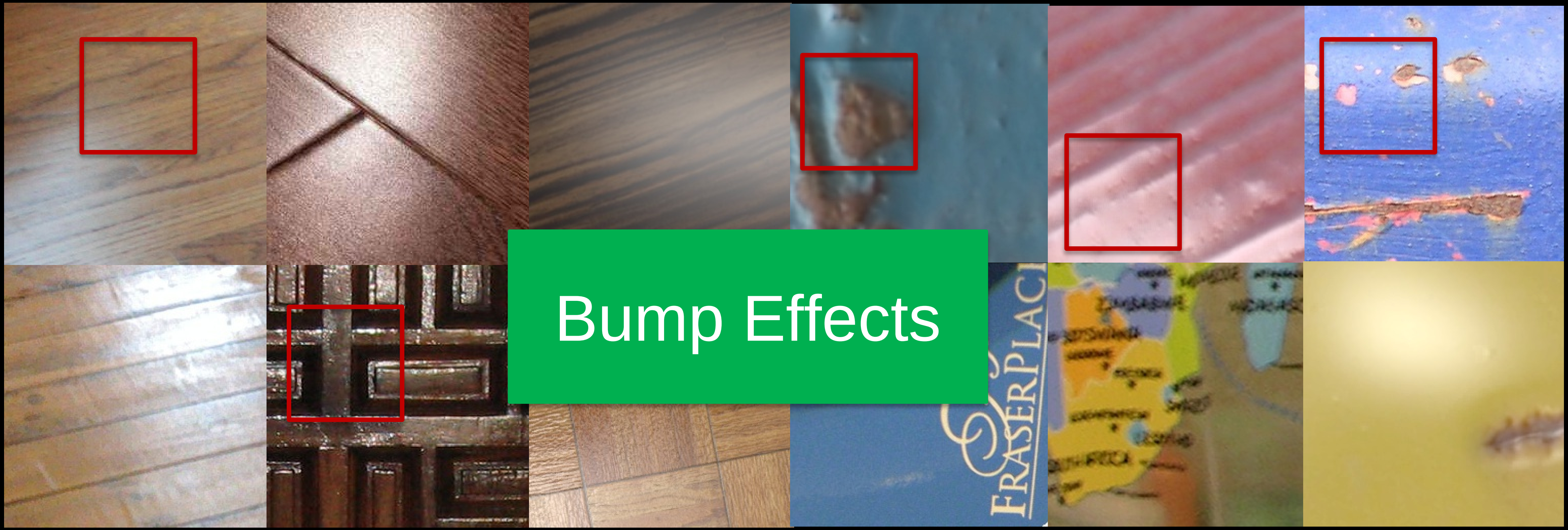
❌ Analysis by Synthesis
- Few SVBRDF data available
- Non-trivial to sample SVBRDF space



Homogeneous Specular
(Ward Model)

Unlabeled Image Contains Information

Unlabeled Image Contains Information



Bump Effects

Key Observation

Forward Mapping: Appearance Modeling

$$A = f(I)$$

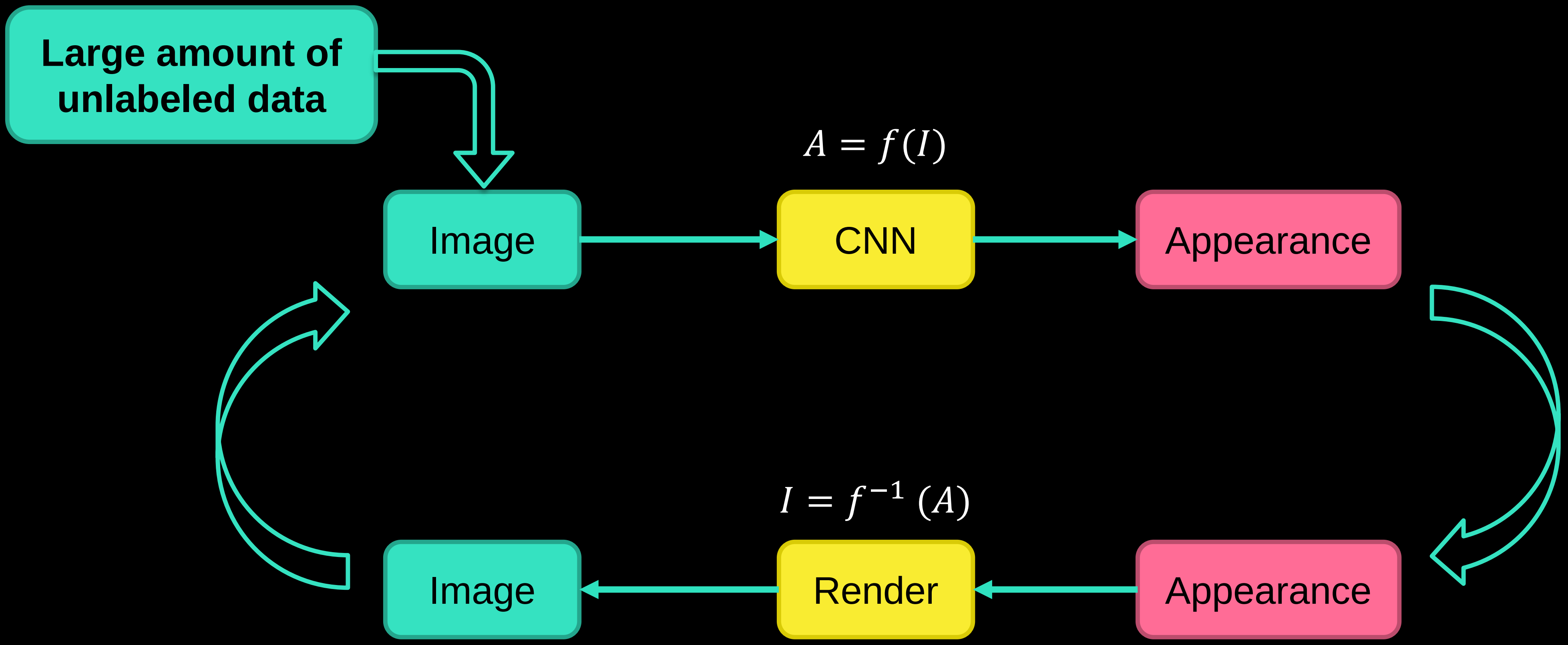
Image → CNN → Appearance

Inverse Mapping: Render

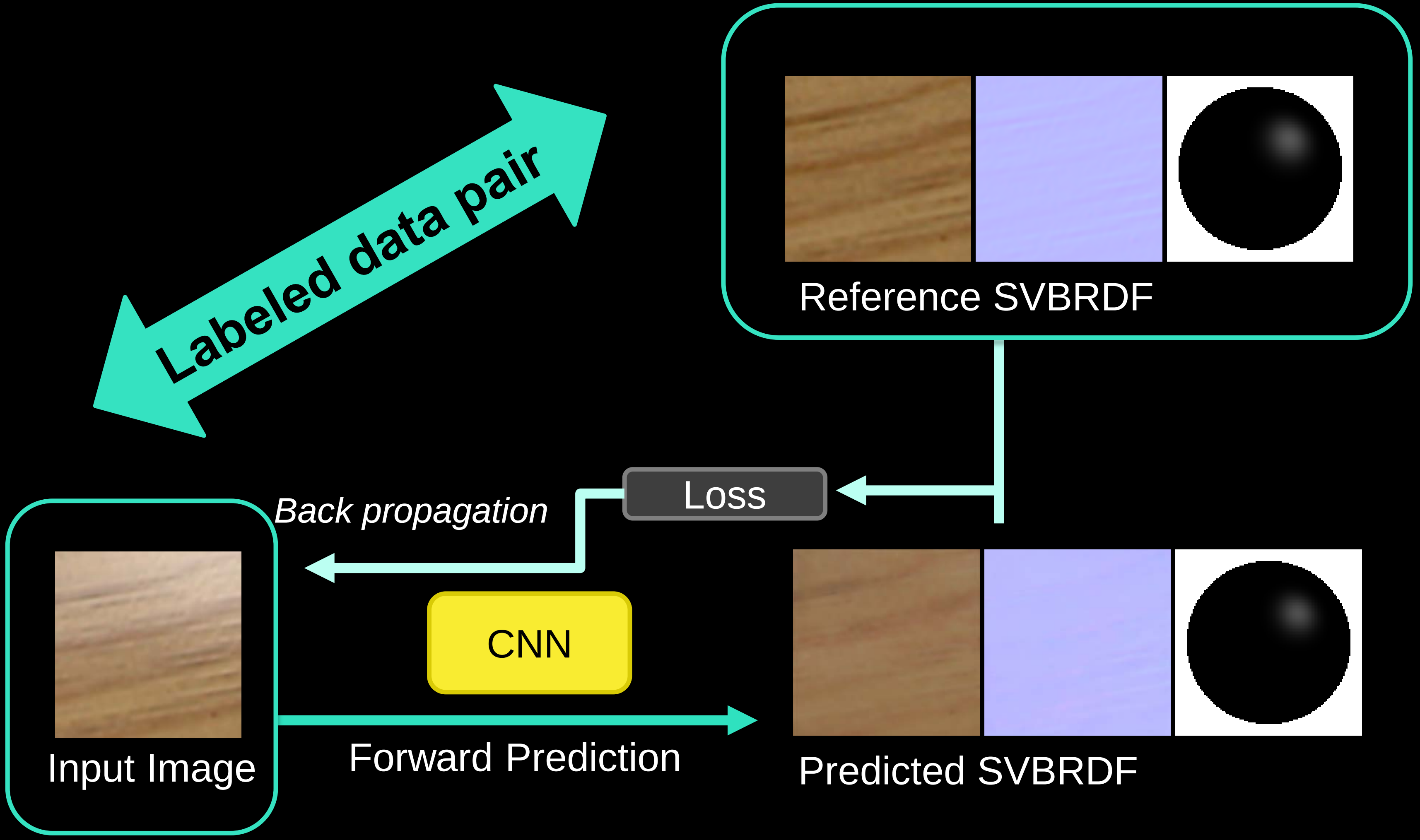
$$I = f^{-1}(A)$$

Image ← Render ← Appearance

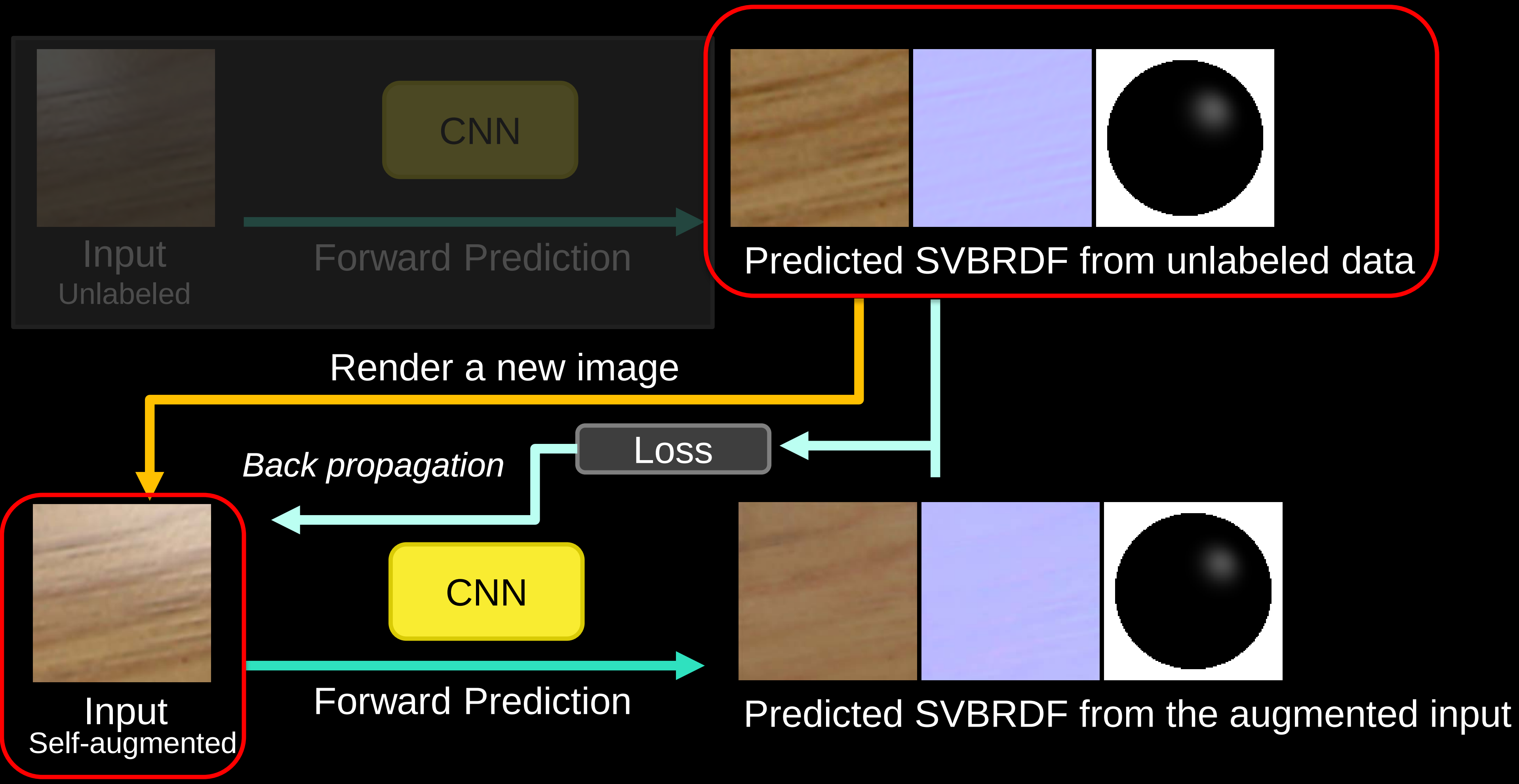
Key Observation



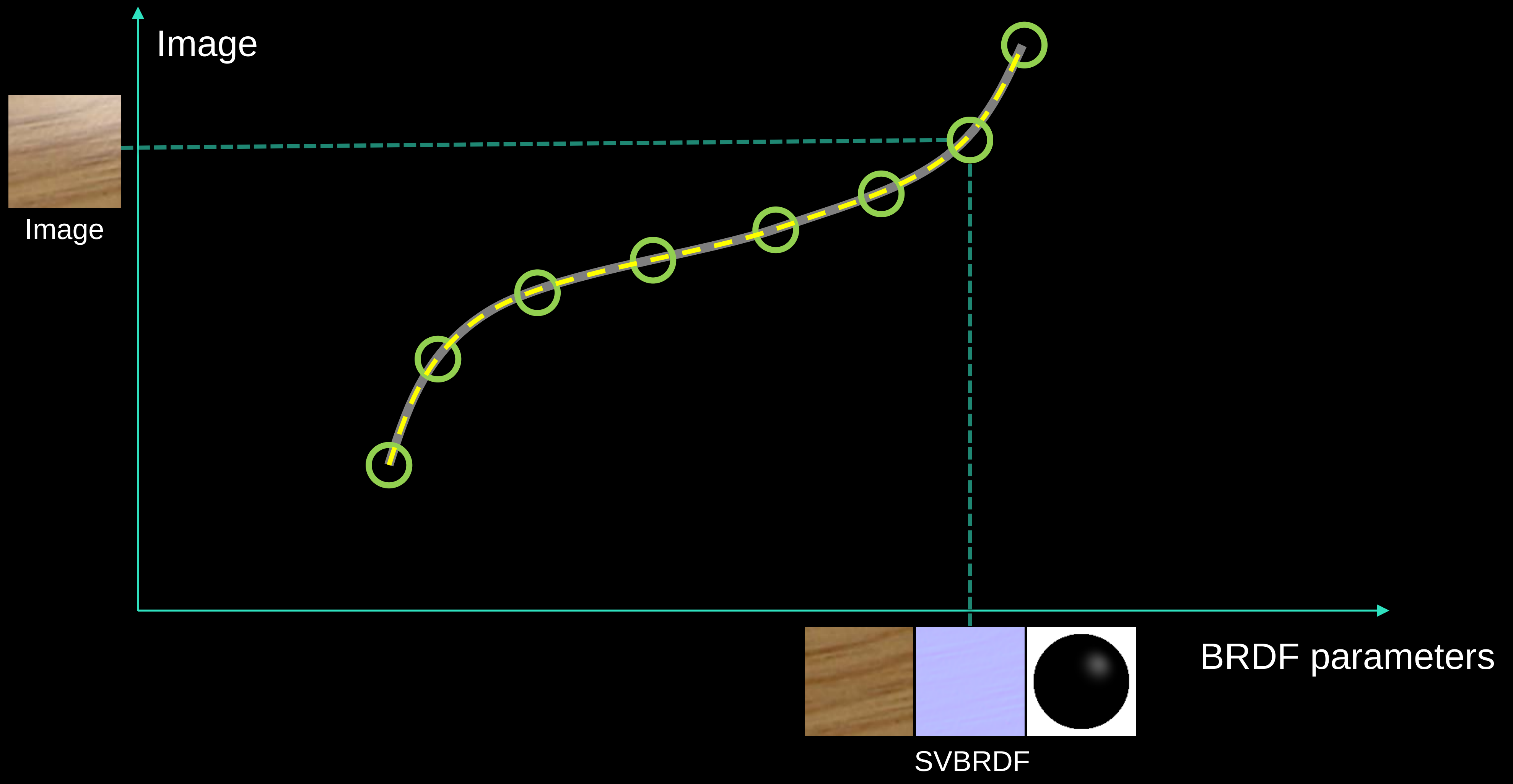
Self-Augmented Training



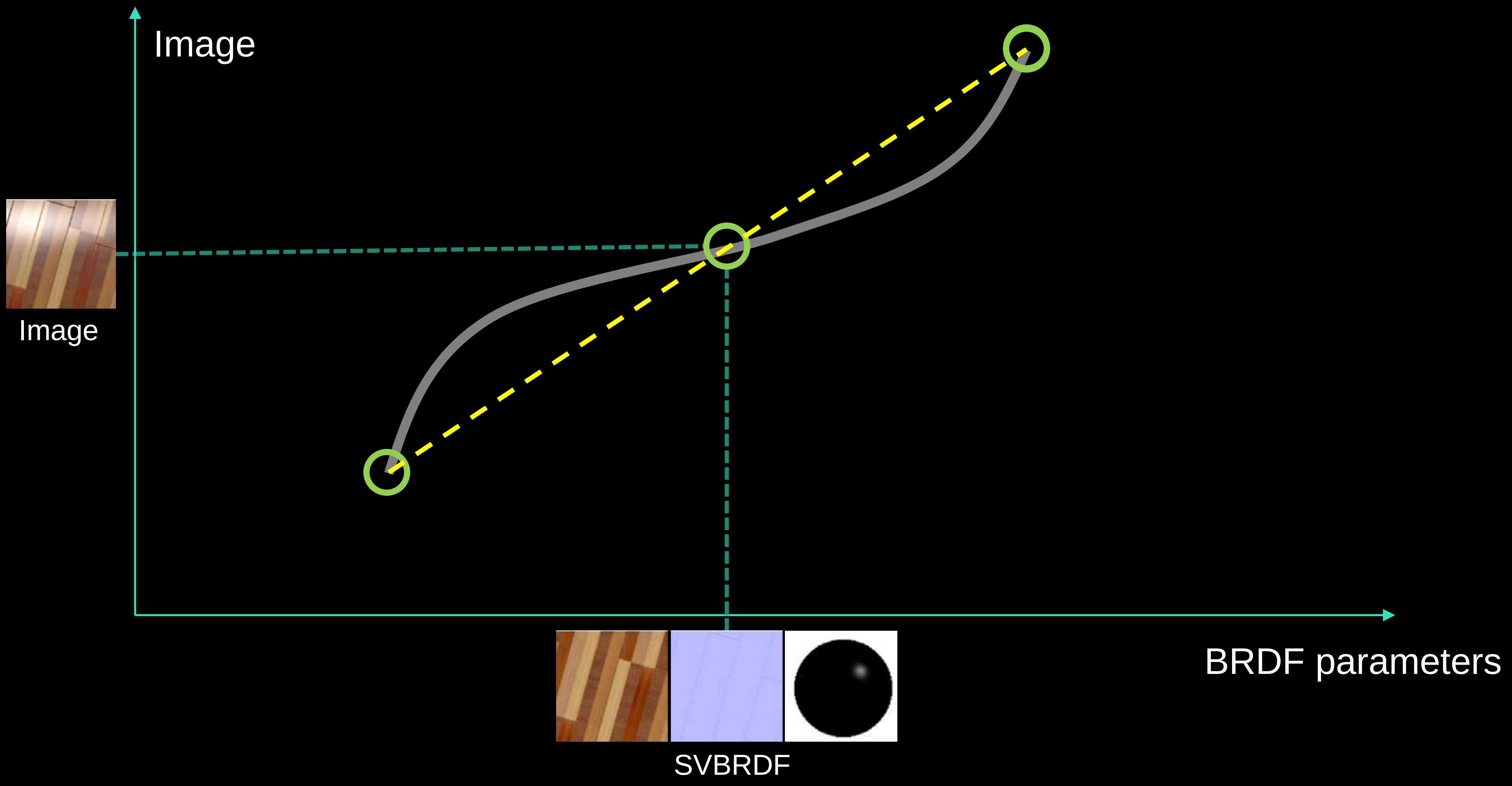
Self-Augmented Training



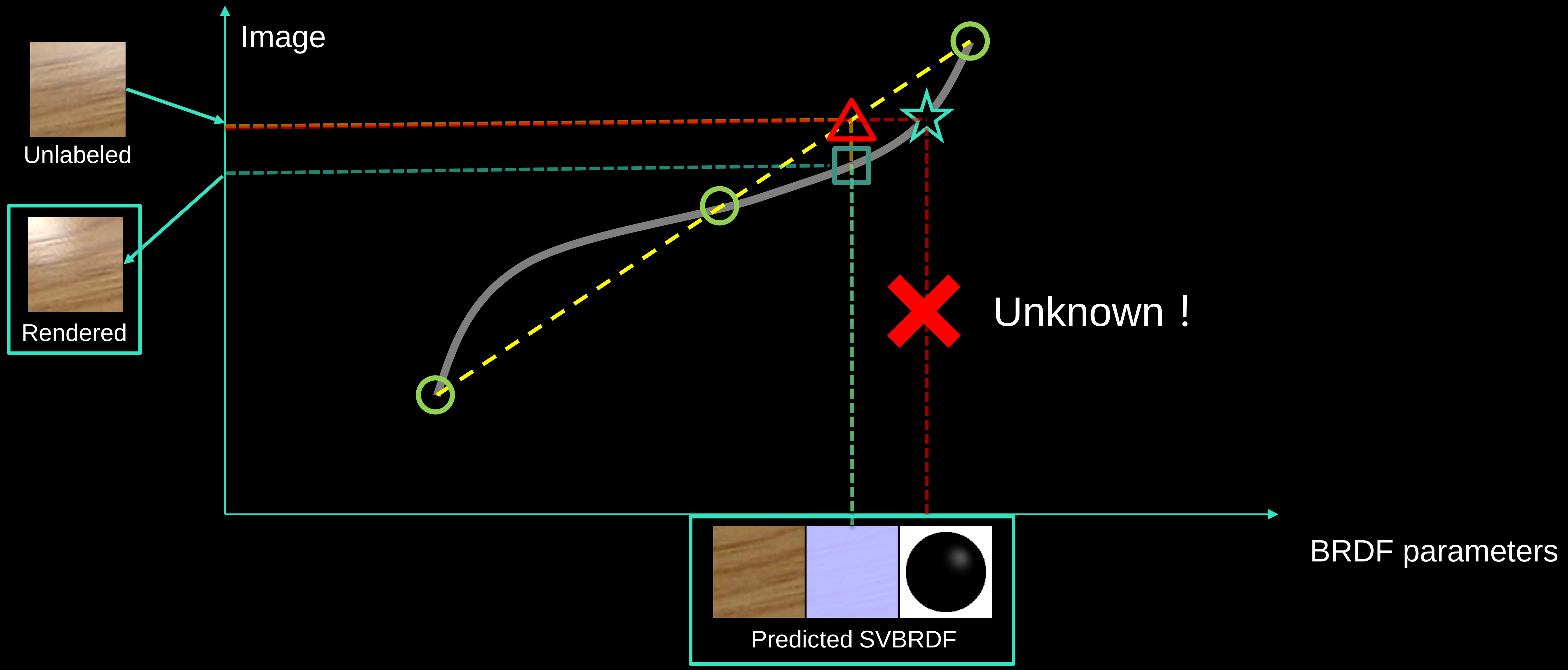
1D illustration of Self-Augmentation



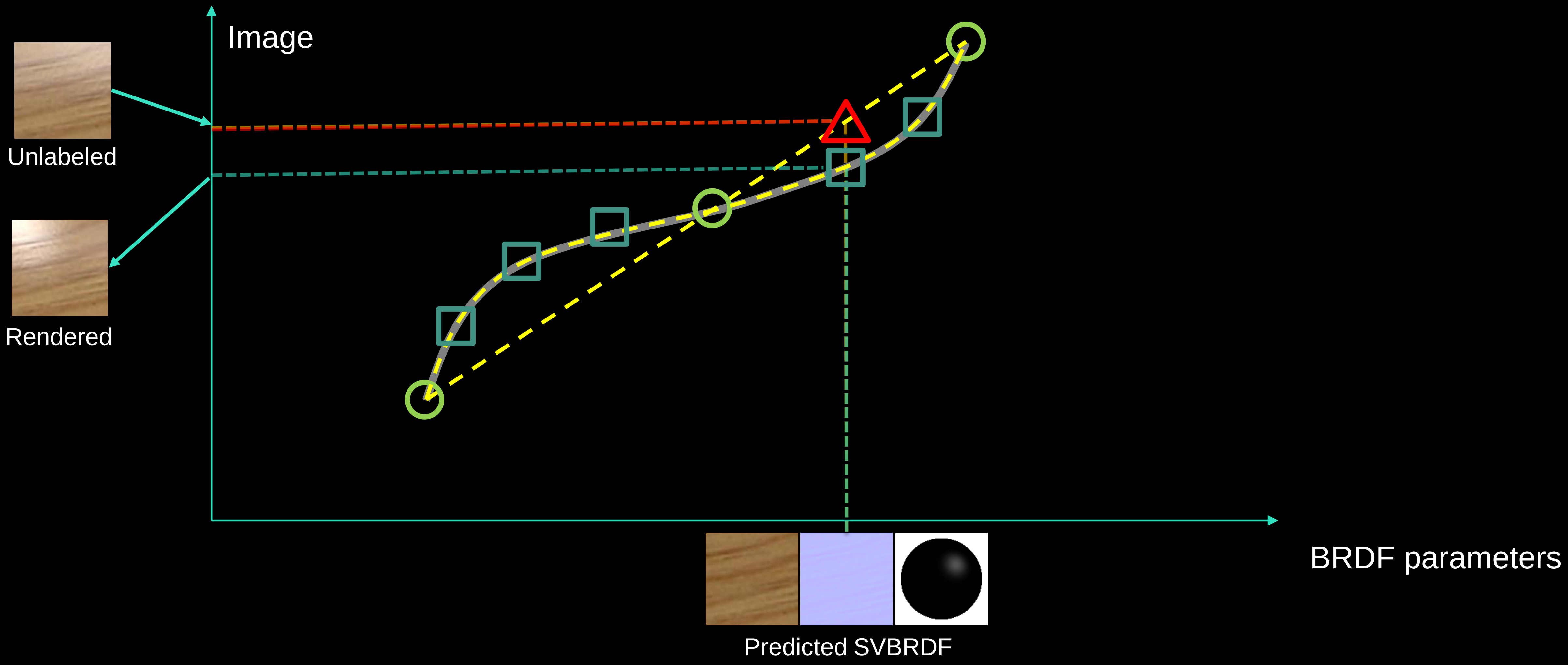
1D illustration of Self-Augmentation



1D illustration of Self-Augmentation

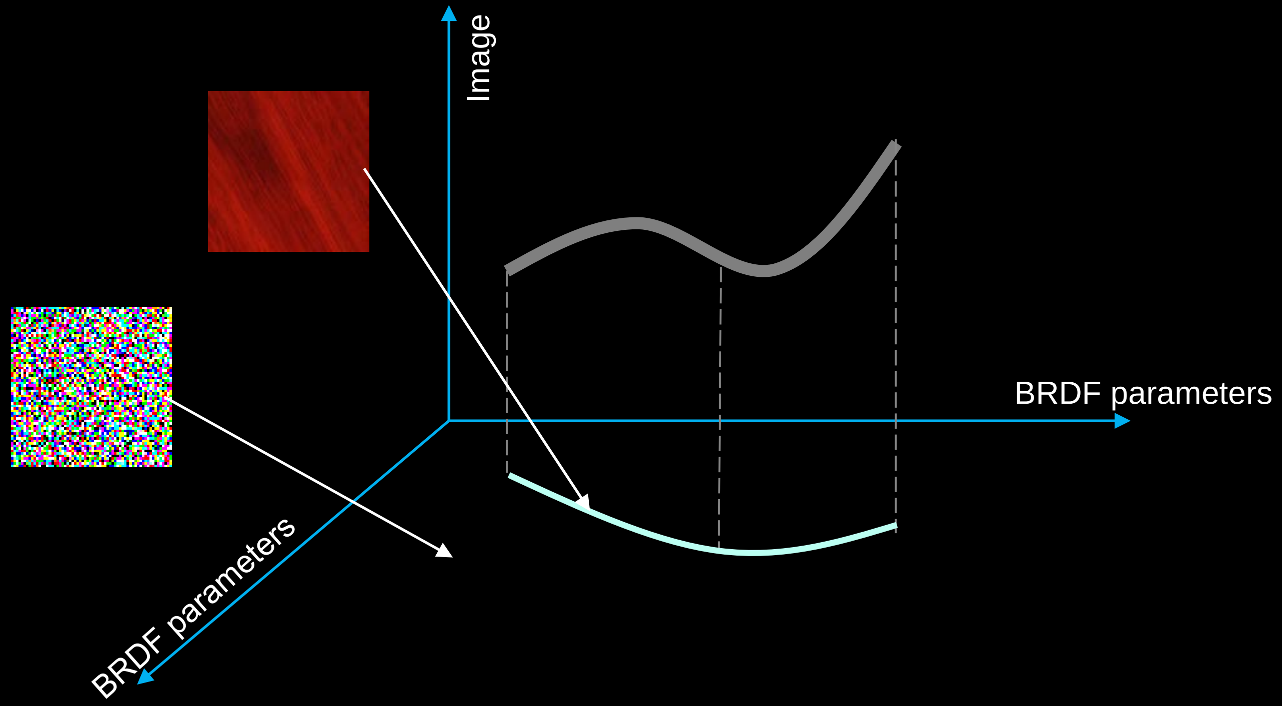


1D illustration of Self-Augmentation



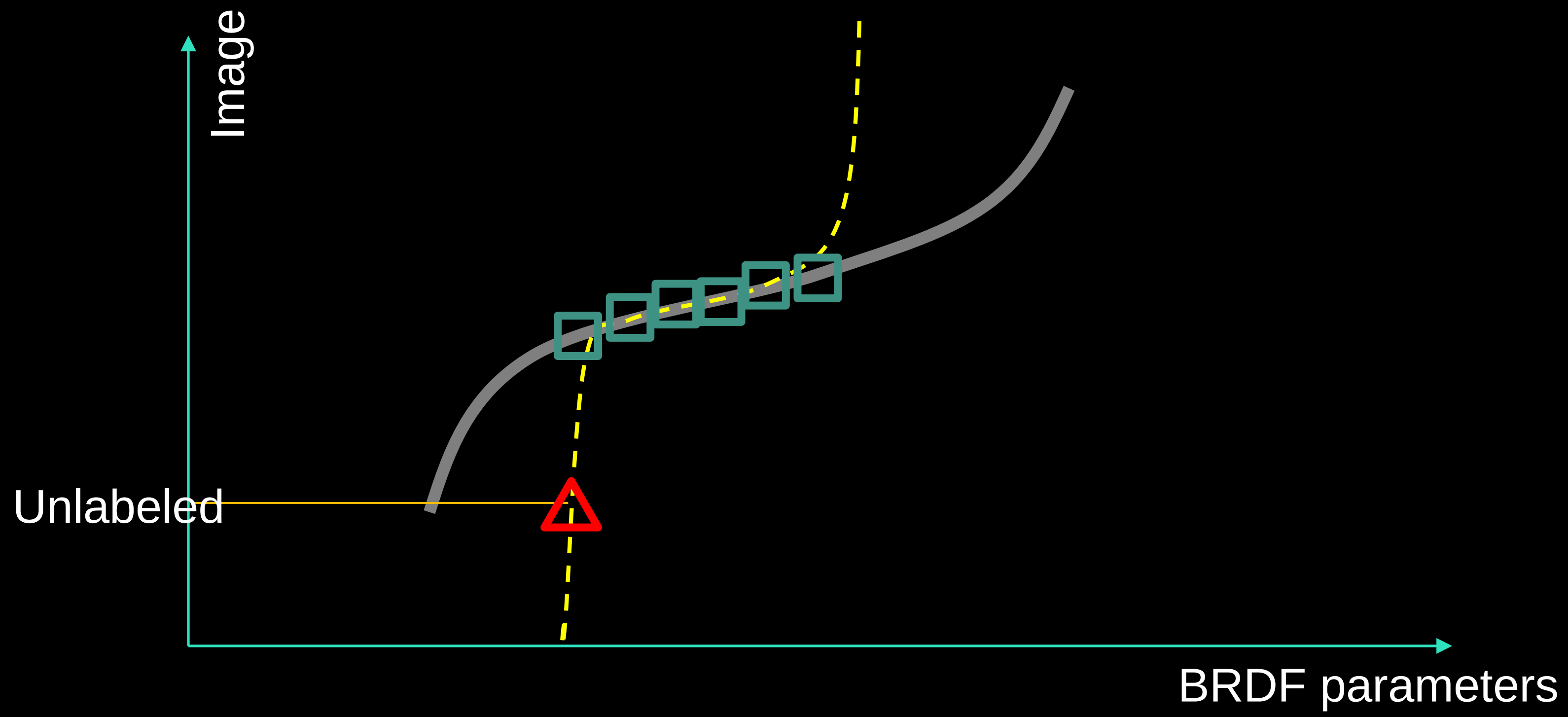
Why SA scheme works

- Exploring the meaningful domain
- Defined in high dimensional space



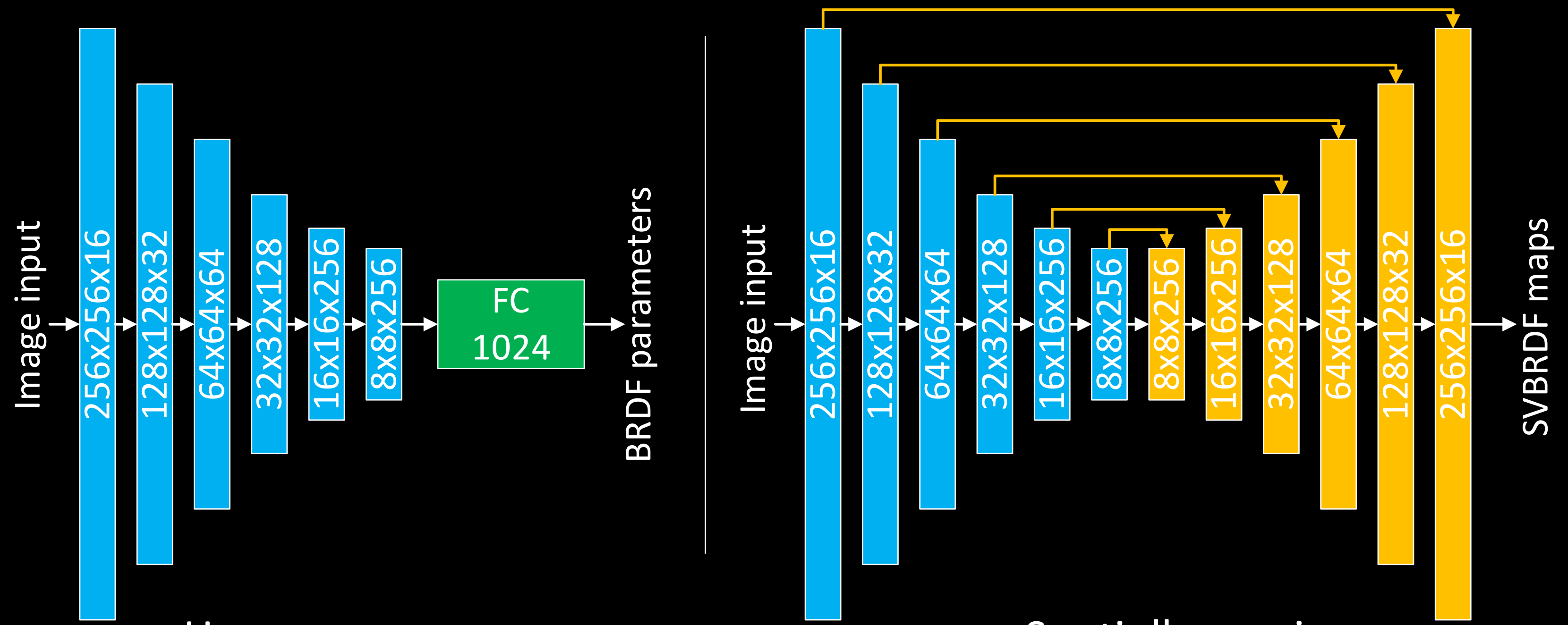
The pitfall

- Local minimal / model collapsing
  - Interleave labeled & SA training minibatches



Network Structure

Fully Convolutional, U-Net



Homogeneous
reflectance parameter estimation

Spatially varying
reflectance parameter estimation

Legend		
	Width x Height x Channel	Convolution (3x3 kernel, stride 2) + BN + ReLU
	Width x Height x Channel	Bi-linear upsample + Convolution (3x3 kernel) + BN + ReLU
	Hidden unit	Fully connection layer

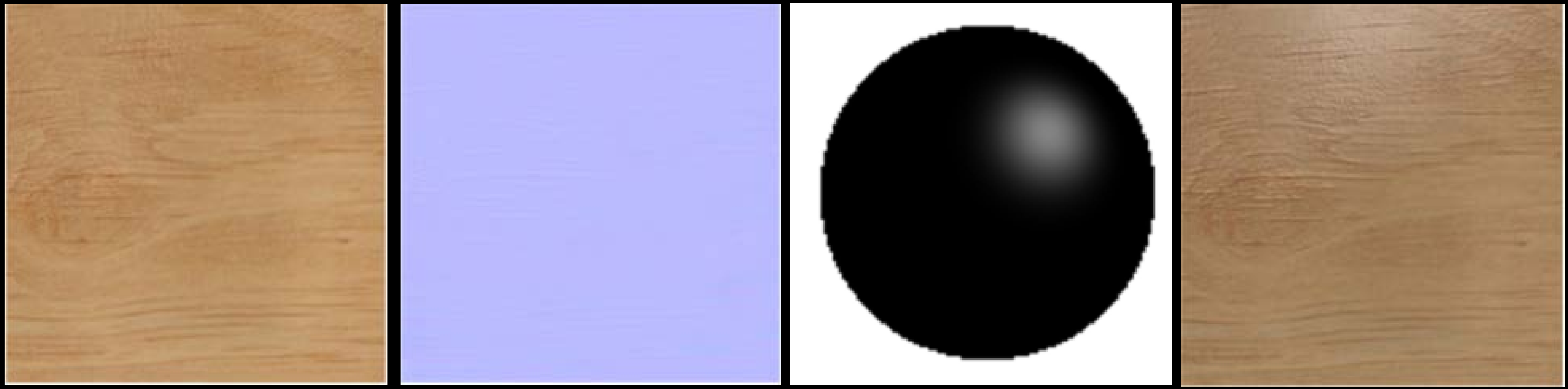
Training Details

- Training data
  - Wood / Metal / Plastic
  - 60 labeled SVBRDFs
  - 1000+ unlabeled photos
  - 256*256 patch
- Performance (Titan X)
  - Training: ~40 hours
  - Inference: ~0.3 sec.

Data and Source Code:
http://msraig.info/~sanet/sanet.htm

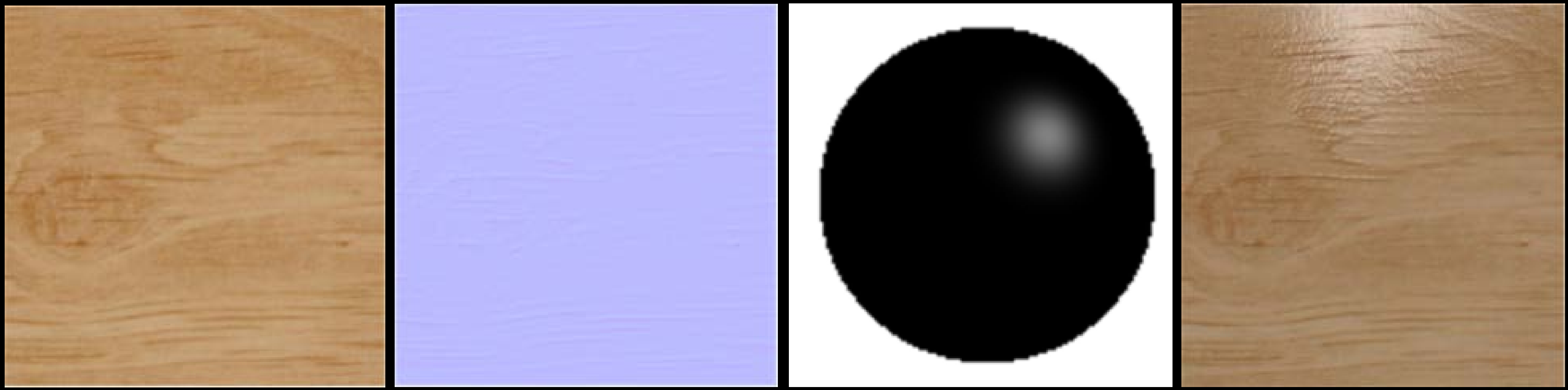
Results - WOOD

Input

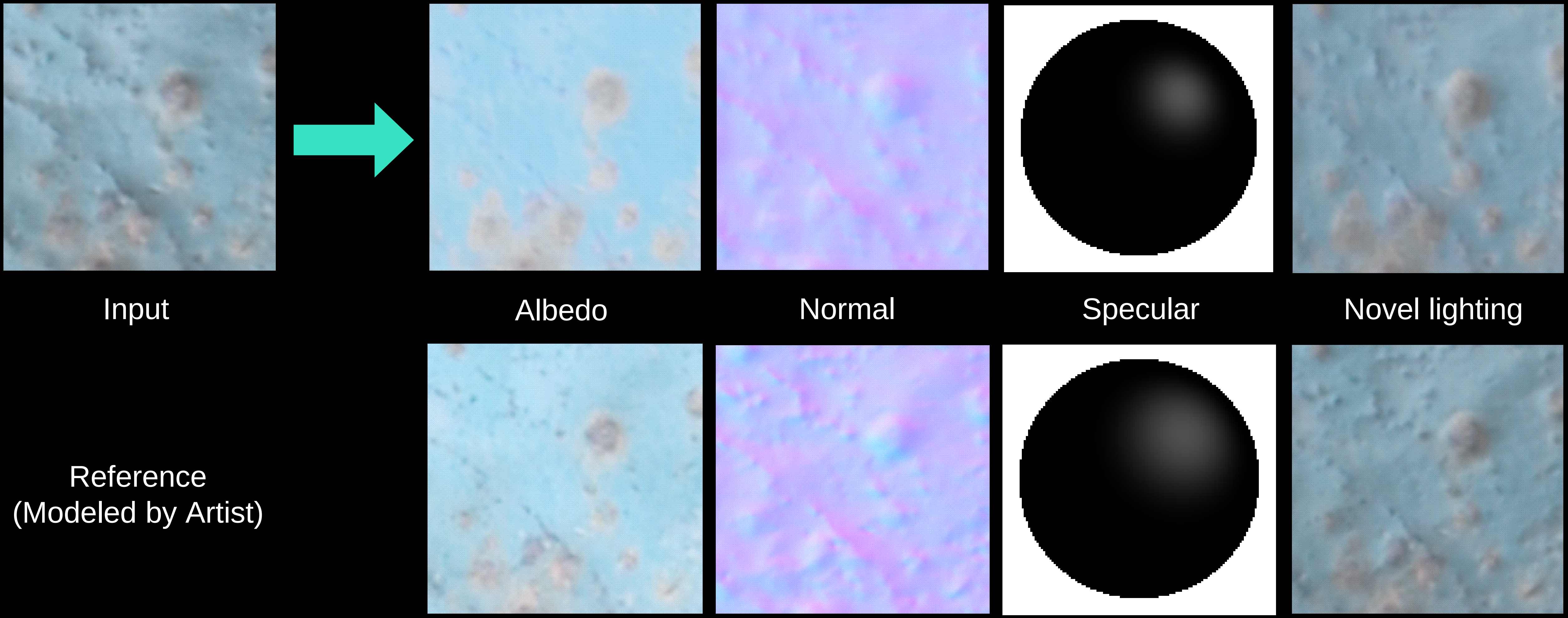


Albedo | Normal | Specular | Novel lighting

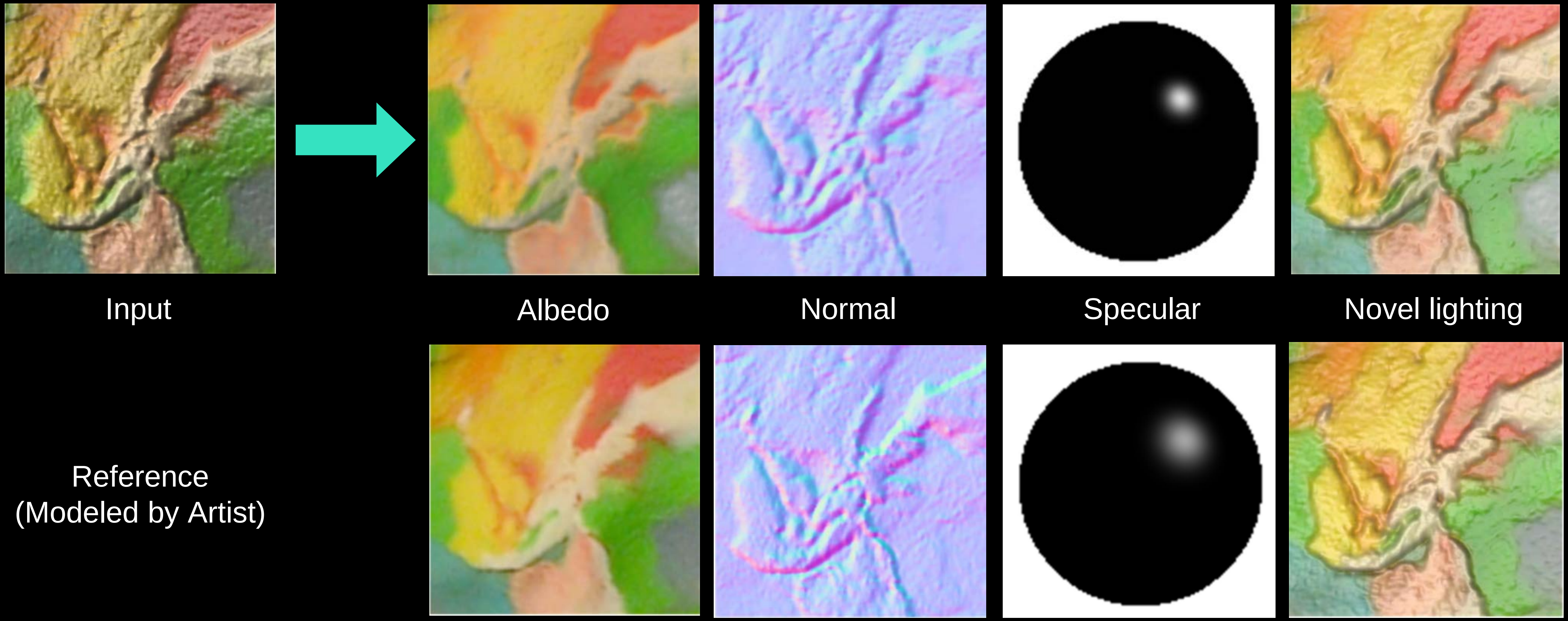
Reference
(Modeled by Artist)



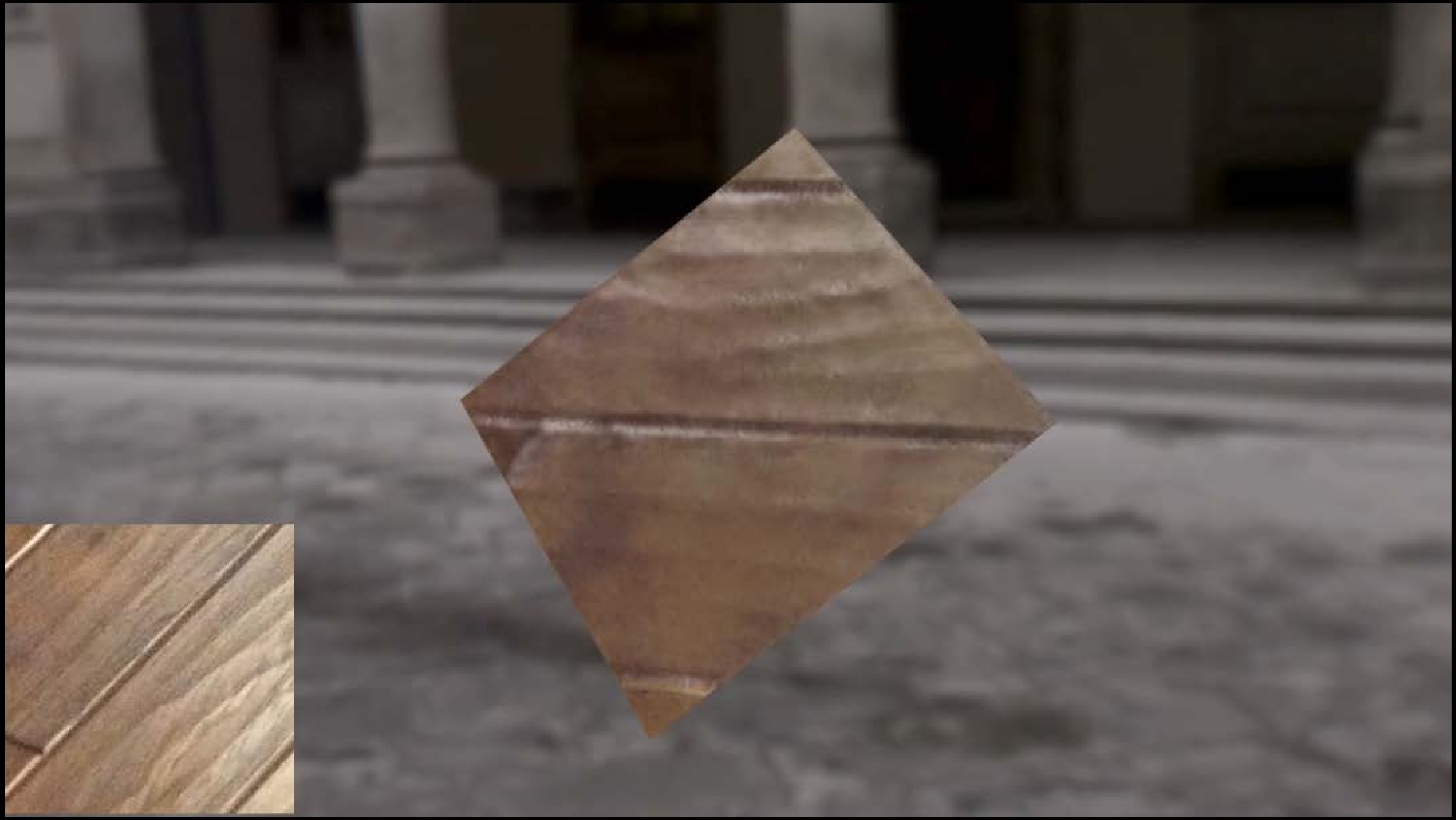
Results - METAL



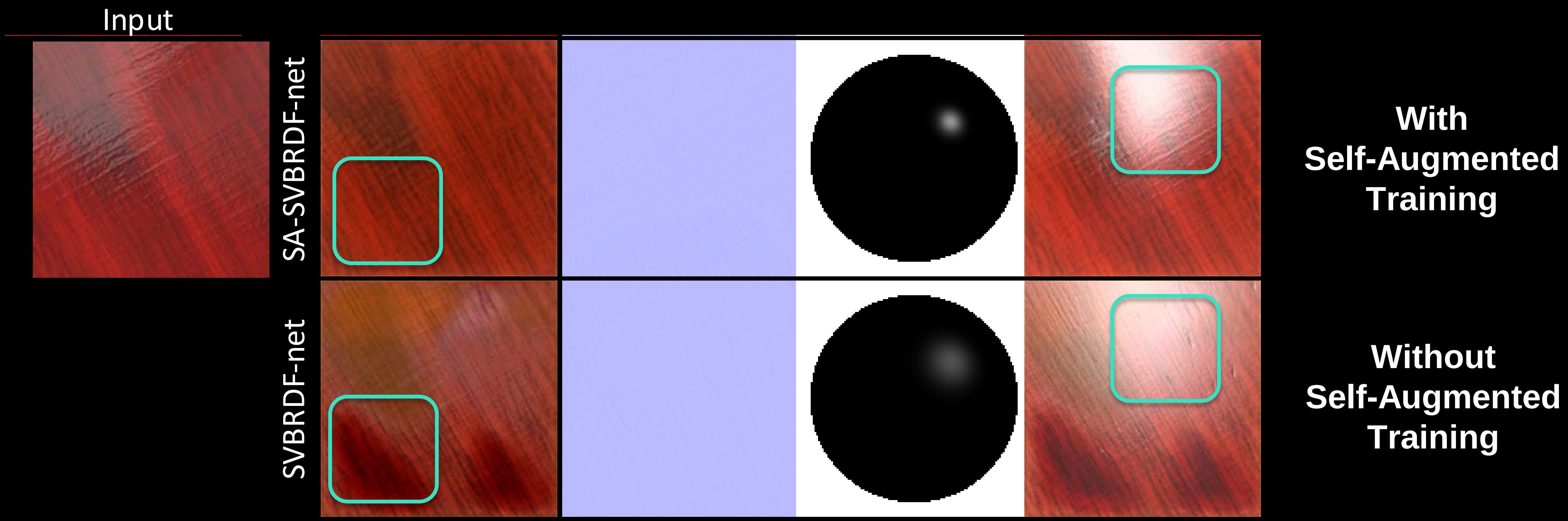
Results - PLASTIC



Relighting Video

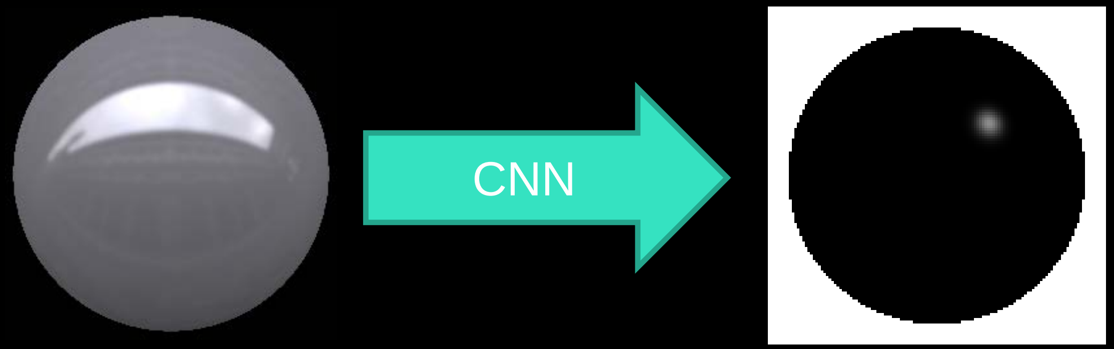


Benefit of Self-Augmentation



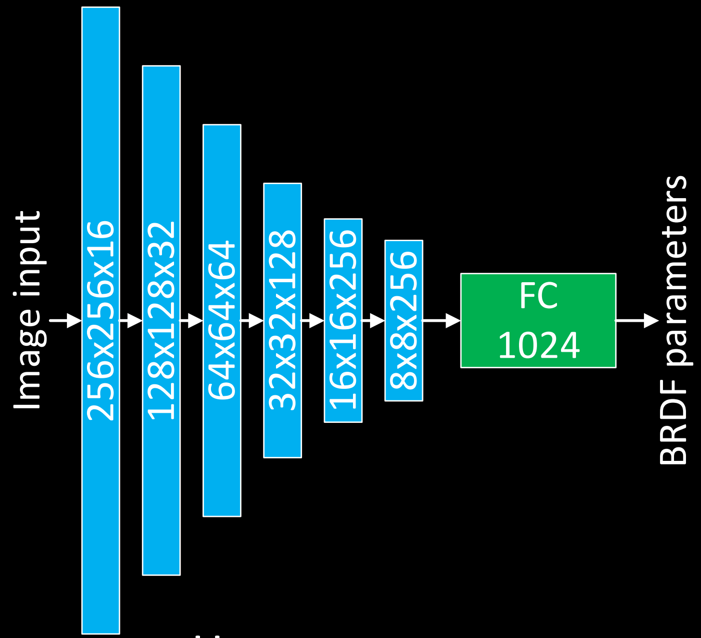
BRDF-Net

- Manageable scale problem for better understanding



Homogeneous BRDF
Rendered on a **sphere**
Under **environment lighting**

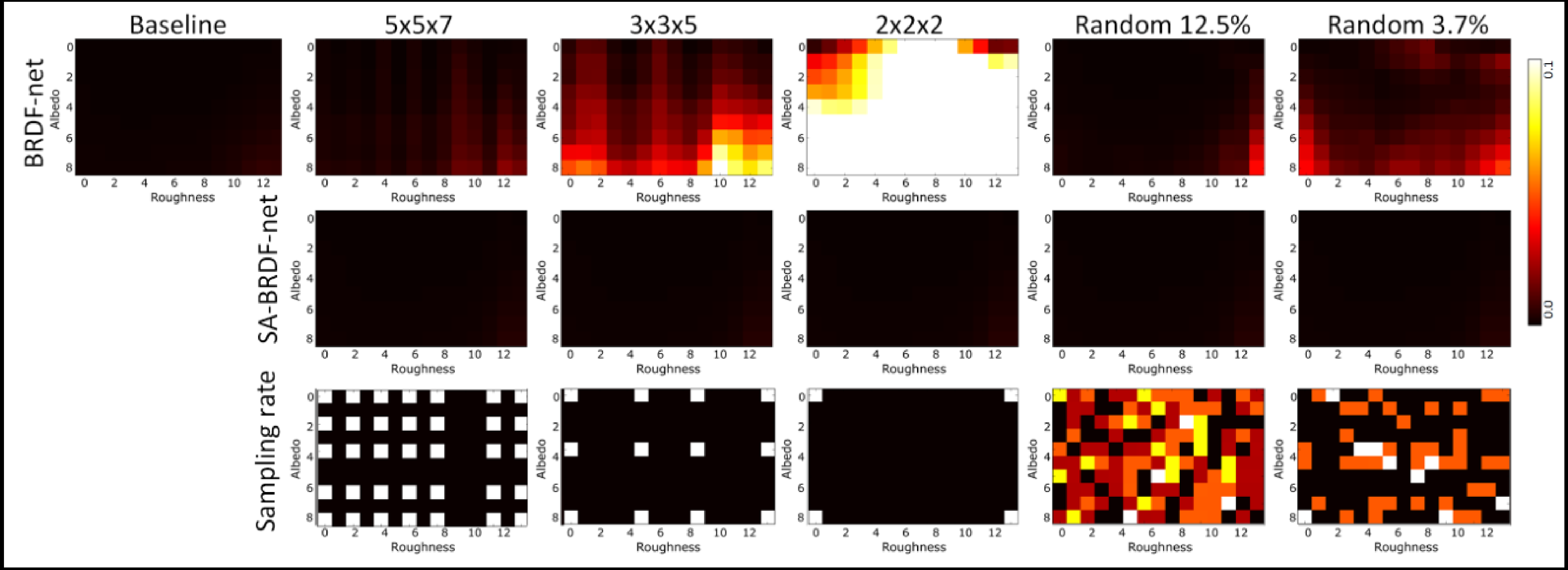
Reflectance parameters



Homogeneous
reflectance parameter estimation

BRDF-Net

- Effects of self-augmentation
- Full labeled data vs Sparse labeled + unlabeled (rest)



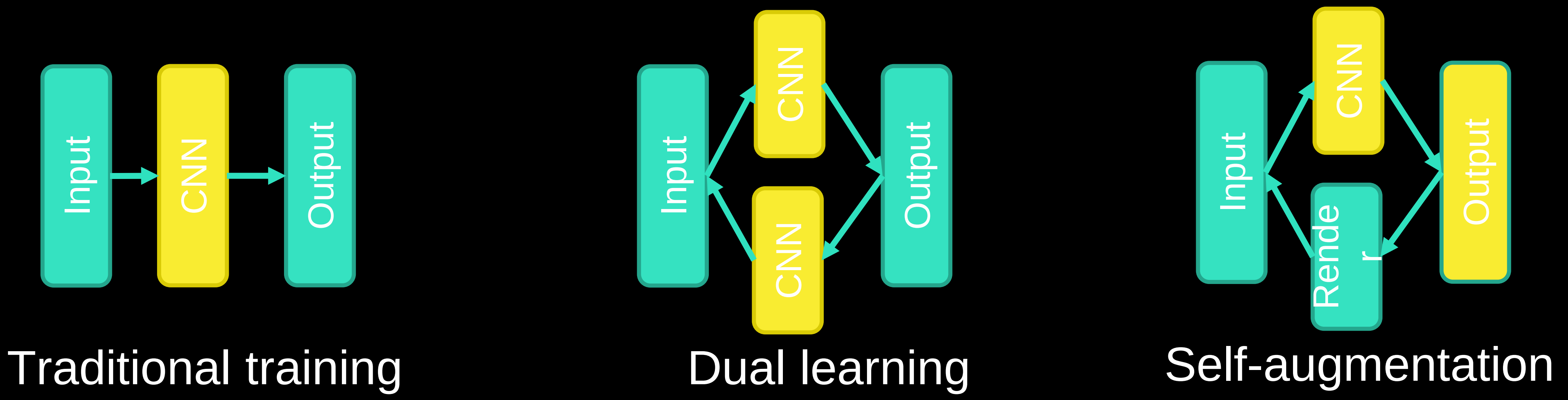
BRDF-Net

- Effects of unlabeled data

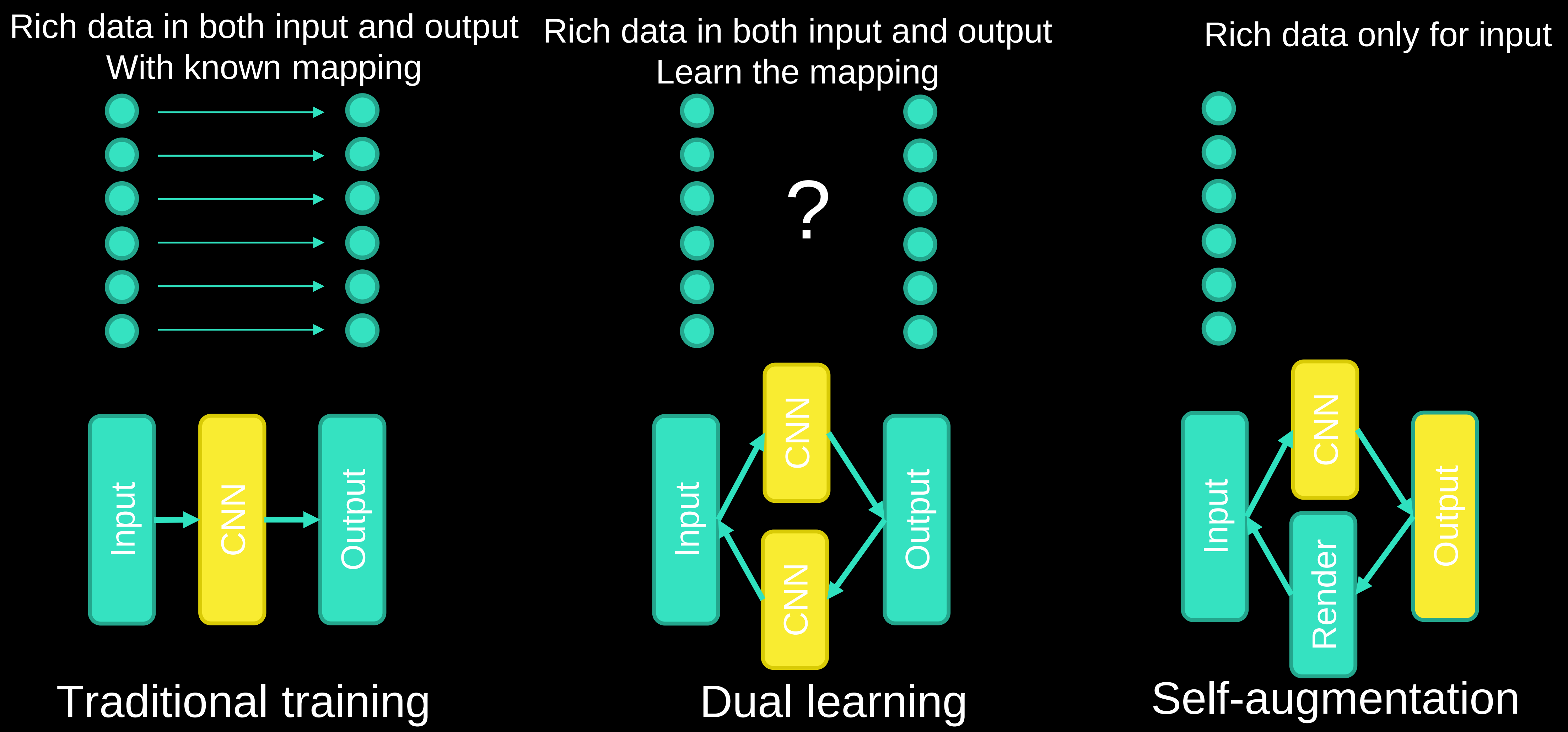
Percent. Labeled	No Self-augmentation	Percentage Unlabeled								
		5	10	20	30	50	70	80	90	95
5	0.002549	0.001395	0.001141	0.000884	0.000689	0.000704	0.000651	0.000578	0.000592	0.000628
10	0.001252	0.001382	0.001027	0.000720	0.000760	0.000671	0.000584	0.000634	0.000592	
20	0.000746	0.001155	0.000845	0.000751	0.000621	0.000619	0.000641	0.000513		
30	0.000662	0.000714	0.000648	0.000694	0.000492	0.000548	0.000535			
50	0.000562	0.000660	0.000559	0.000552	0.000506	0.000470				
70	0.000619	0.000601	0.000462	0.000550	0.000499					
80	0.000553	0.000542	0.000421	0.000413						
90	0.000546	0.000505	0.000471							
95	0.000550	0.000471								
100	0.000499									

Self-augmentation vs dual learning

- Parallel scheme
  - With different known components



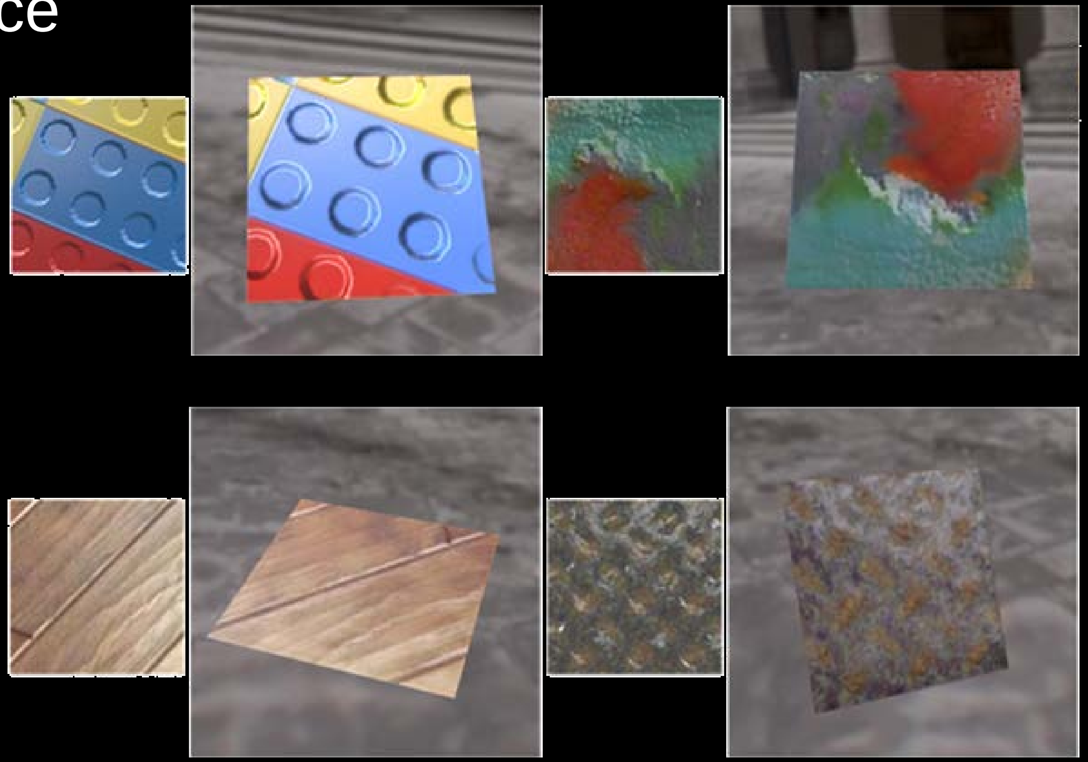
Self-augmentation vs dual learning



Conclusion

- Self-augmented training
  - Single image => Plausible appearance
  - Labeled + Unlabeled training

- Future Work
  - More complex surface appearance
  - Self-augmentation for other tasks



Acknowledgements

- Anonymous Reviewers
- Beijing Film Academy
- NSF grant: IIS-1350323

30 JULY – 3 AUGUST Los Angeles
SIGGRAPH 2017

THANKS!

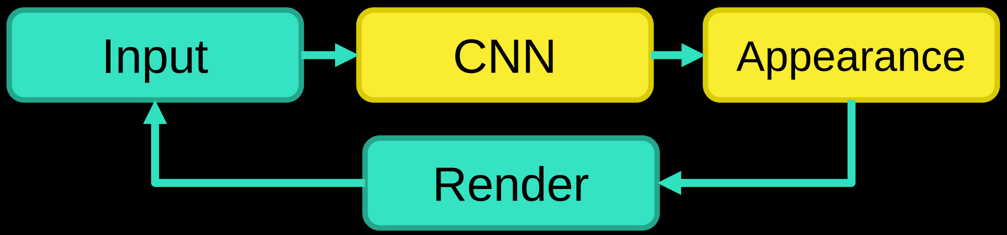
30 JULY – 3 AUGUST Los Angeles
SIGGRAPH 2017

BACKUP SLIDES

Self-augmentation vs dual learning

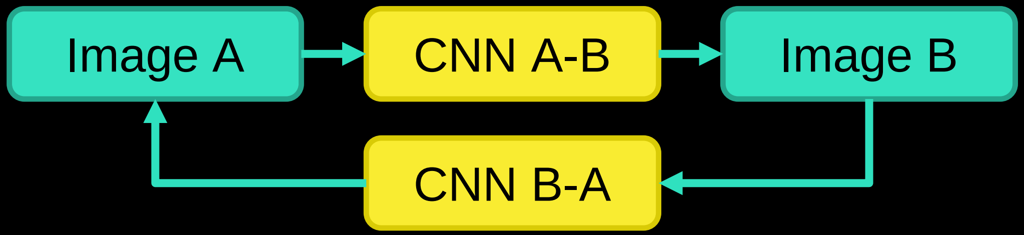
Self-augmentation

- Unlabeled Input
- Known Inverse Mapping



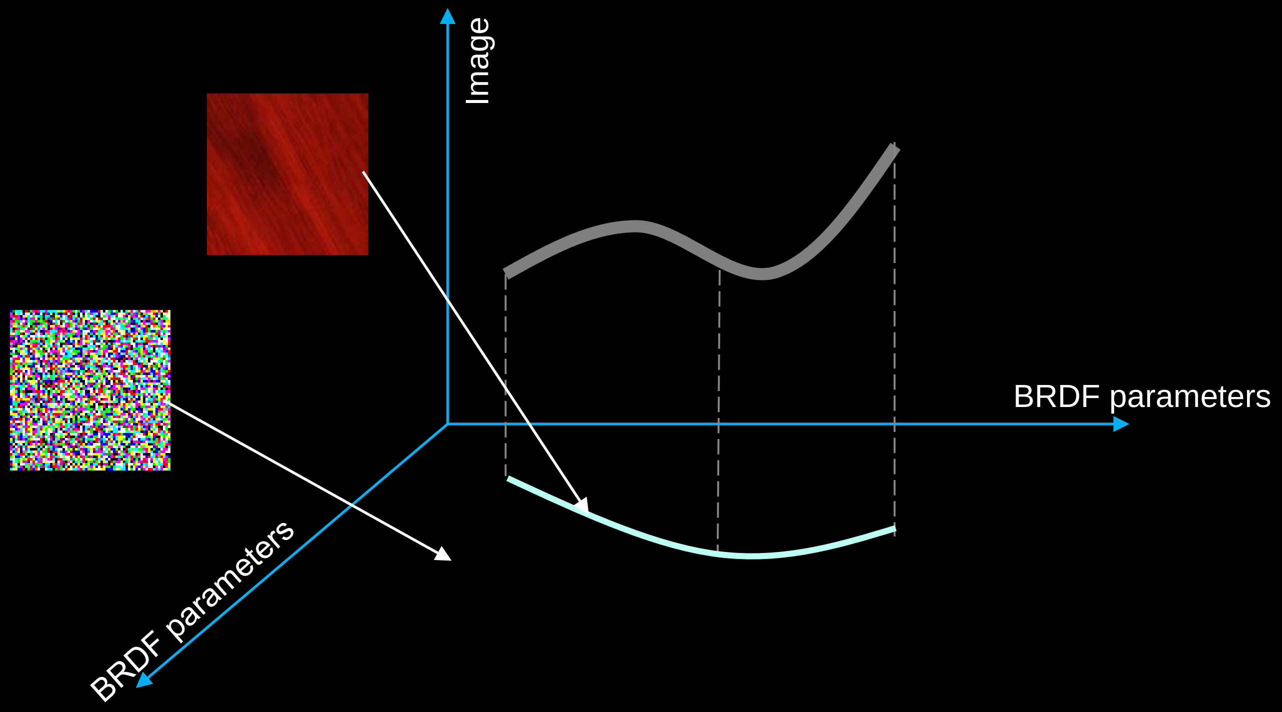
Dual learning

- Unlabeled on two tasks
- Trained dual tasks



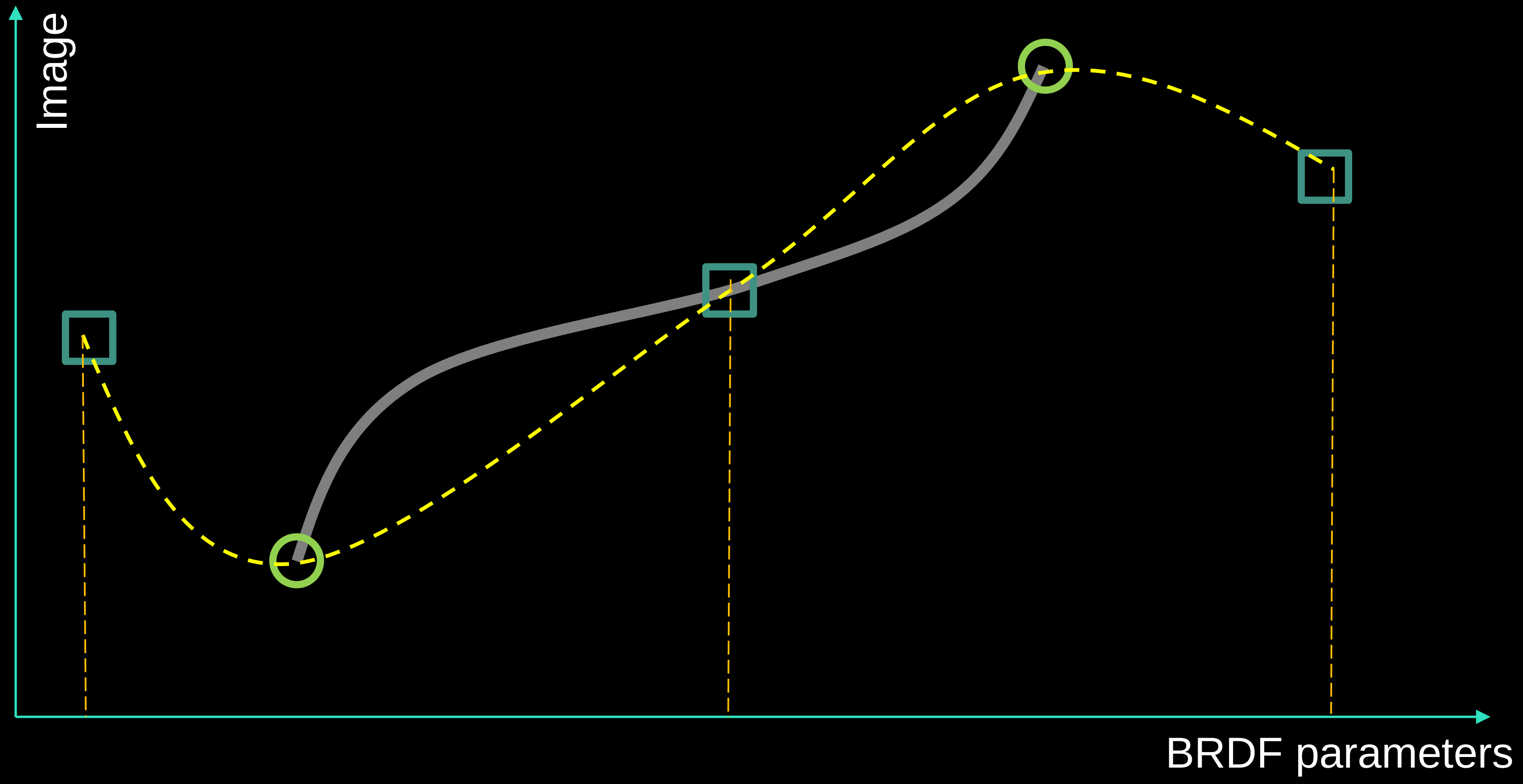
Discussion

- Exploring the meaningful domain defined in high dimensional space



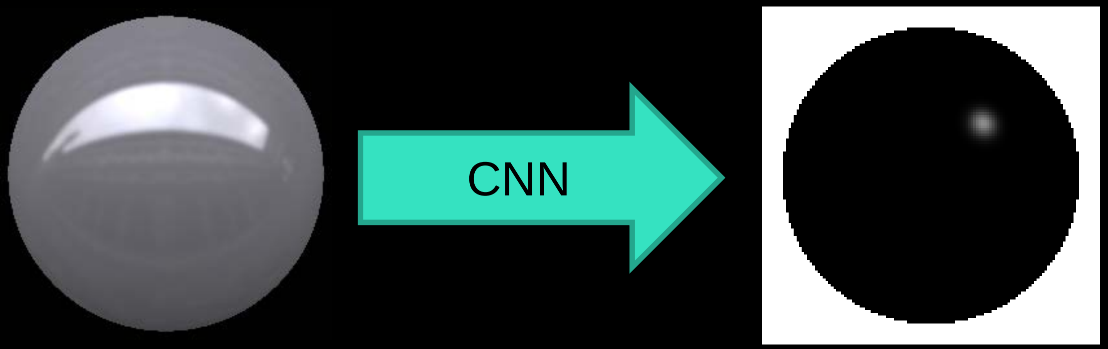
SOLUTION: Self-Augment Training

- An 1D illustration



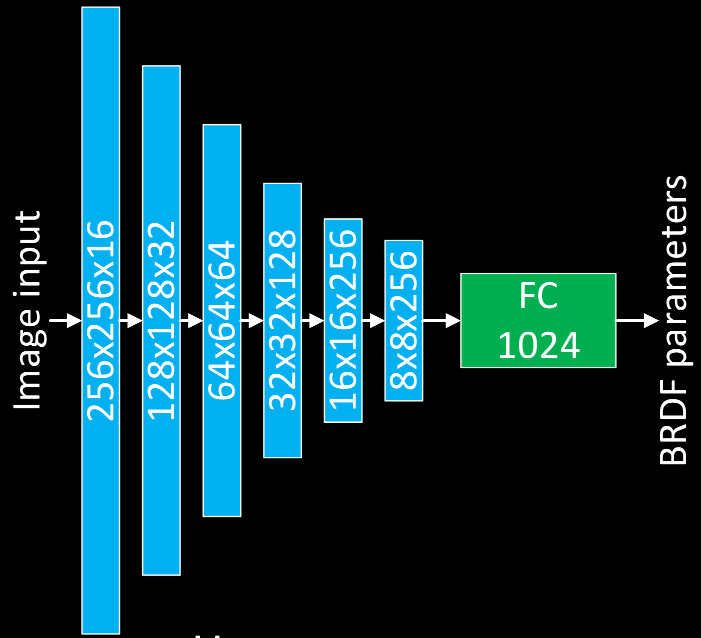
Vaildation: BRDF-NET

- Manageable scale problem for better understanding



Homogeneous BRDF
Rendered on a **sphere**
Under **environment lighting**

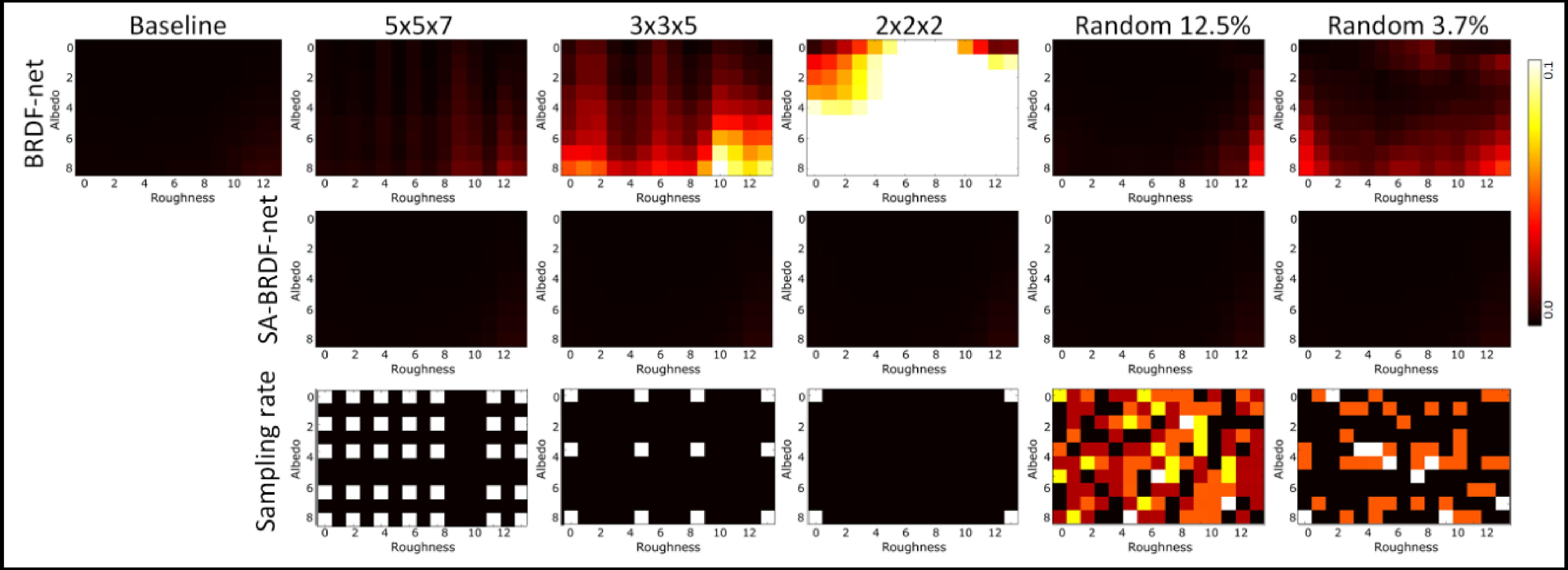
Reflectance parameters



Homogeneous
reflectance parameter estimation

Vaildation: BRDF-NET

- Effects of self-augmentation
  - Full labeled data v.s. sparse labeled + rest data unlabeled



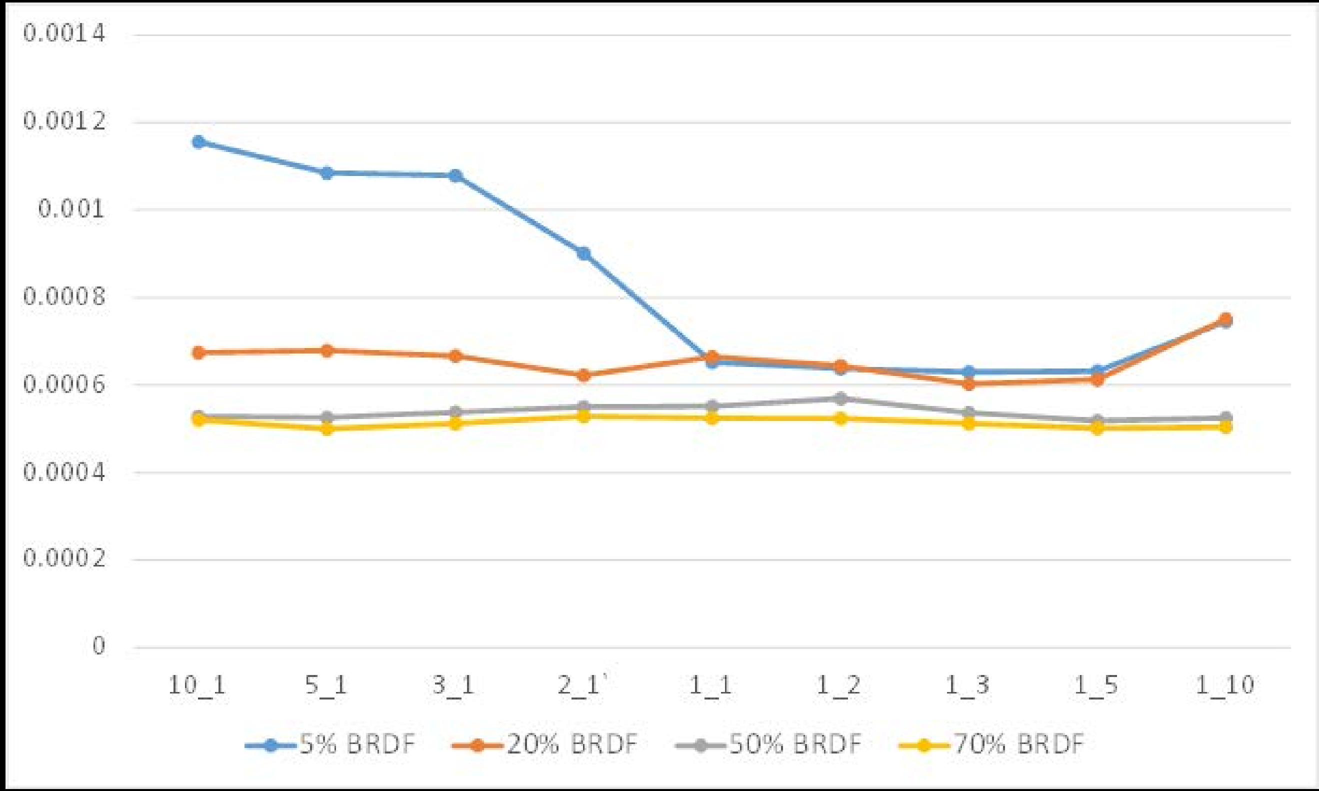
Vaildation: BRDF-NET

- Effects of different amount of unlabeled data

Percent. Labeled	No Self-augmentation	Percentage Unlabeled								
		5	10	20	30	50	70	80	90	95
5	0.002549	0.001395	0.001141	0.000884	0.000689	0.000704	0.000651	0.000578	0.000592	0.000628
10	0.001252	0.001382	0.001027	0.000720	0.000760	0.000671	0.000584	0.000634	0.000592	
20	0.000746	0.001155	0.000845	0.000751	0.000621	0.000619	0.000641	0.000513		
30	0.000662	0.000714	0.000648	0.000694	0.000492	0.000548	0.000535			
50	0.000562	0.000660	0.000559	0.000552	0.000506	0.000470				
70	0.000619	0.000601	0.000462	0.000550	0.000499					
80	0.000553	0.000542	0.000421	0.000413						
90	0.000546	0.000505	0.000471							
95	0.000550	0.000471								
100	0.000499									

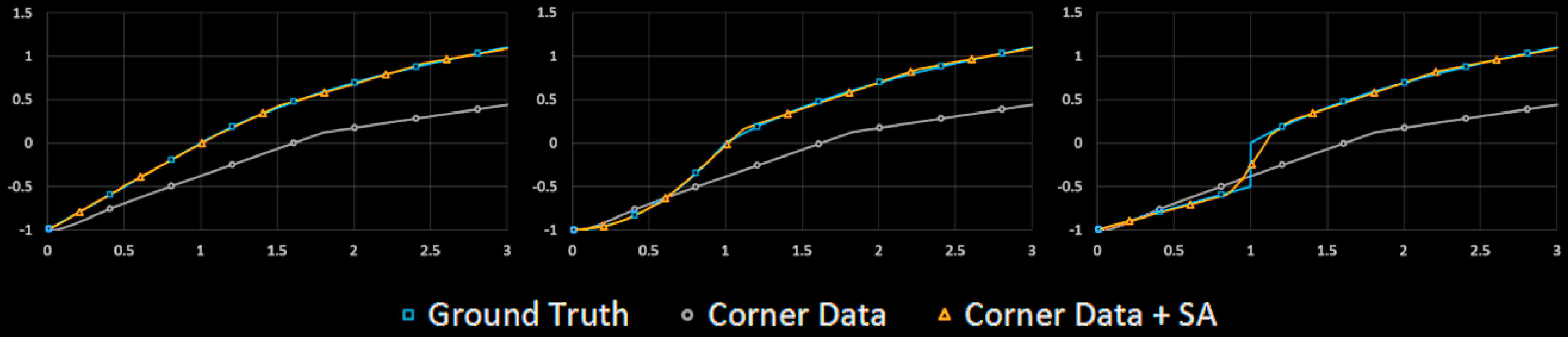
Vaildation: BRDF-NET

- Interleave training ratio between unlabeled / labeled data



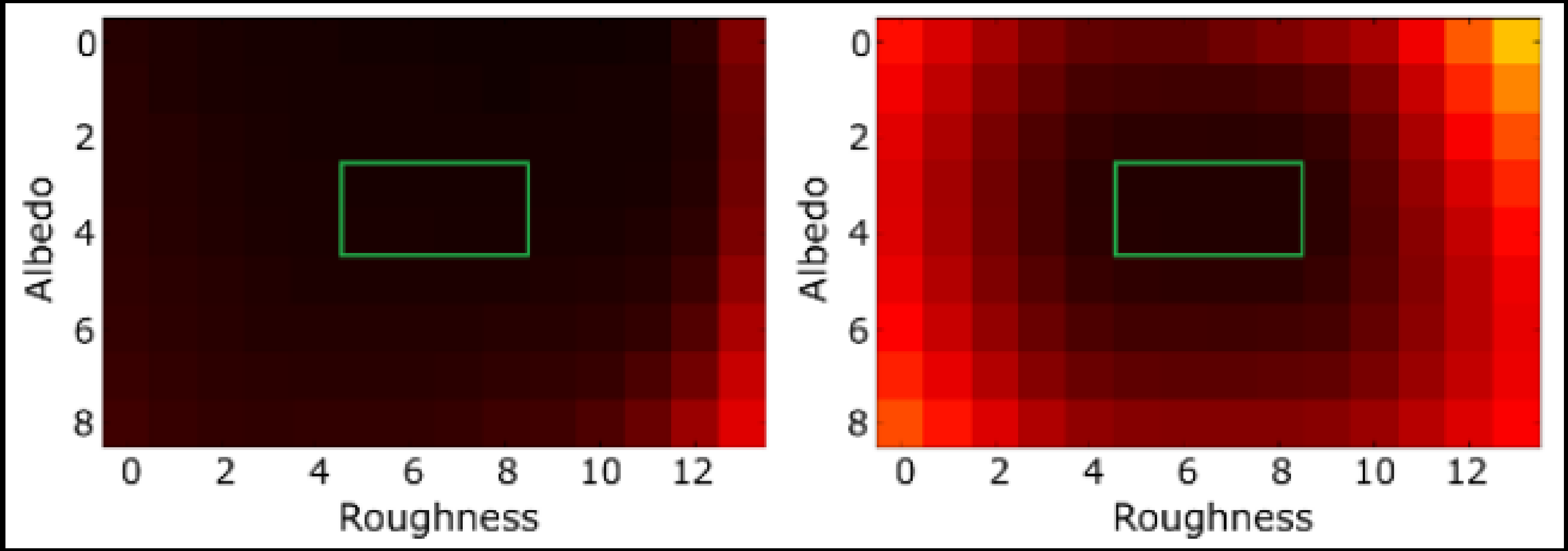
SOLUTION: Self-Augment Training

- An 1D training illustration
  - Regression with 2 layer MLP
  - Only 2 labeled data at the ends / unlabeled data for full range



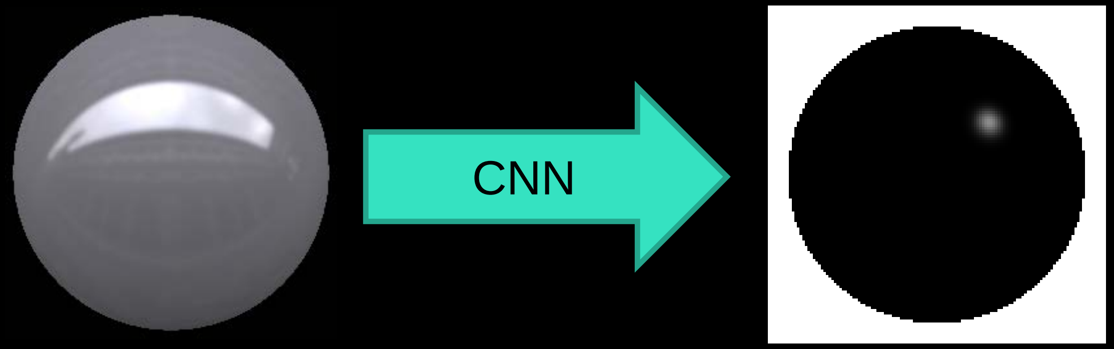
Vaildation: BRDF-NET

- Vaildation on convex hull assumption



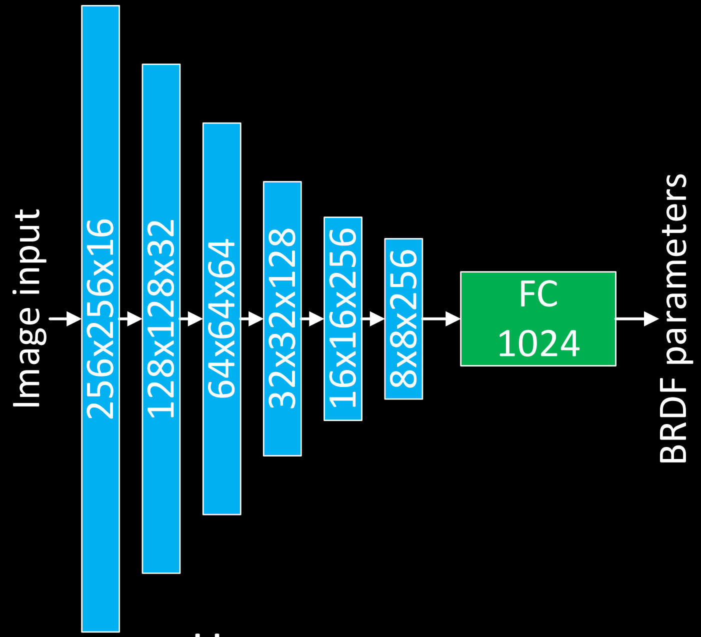
In-Depth Validation of SA scheme

- Convex Hull
  - labeled data should cover whole space
- Interleave ratio
  - 1 : 1



Homogeneous BRDF
Rendered on a **sphere**
Under **environment lighting**

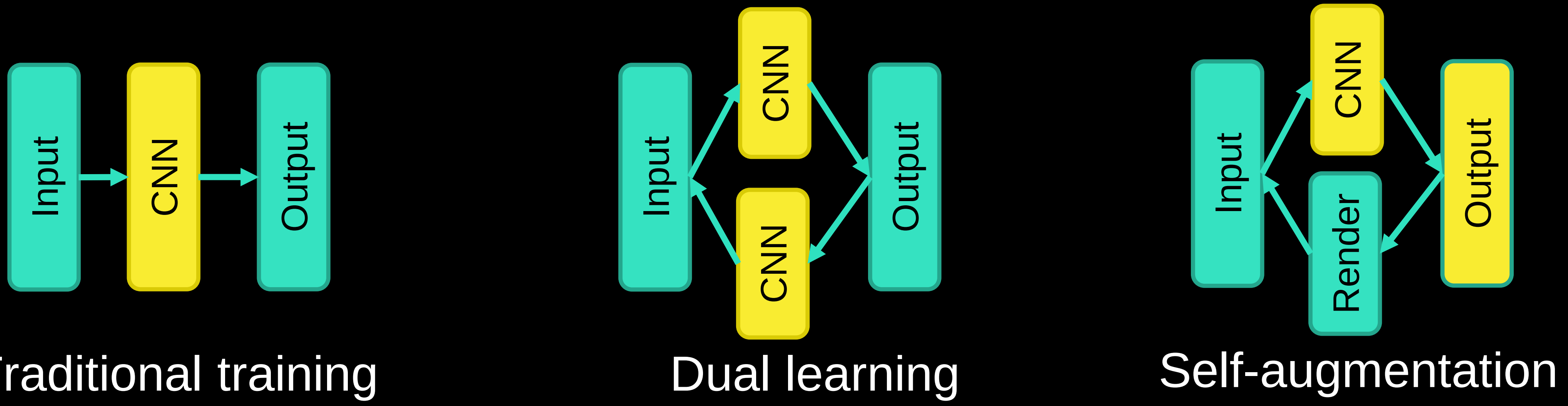
Reflectance parameters



Homogeneous
reflectance parameter estimation

Self-augmentation vs dual learning

- Unlabeled Input
- Known Inverse Mapping
- Unlabeled on two tasks
- Trained dual tasks

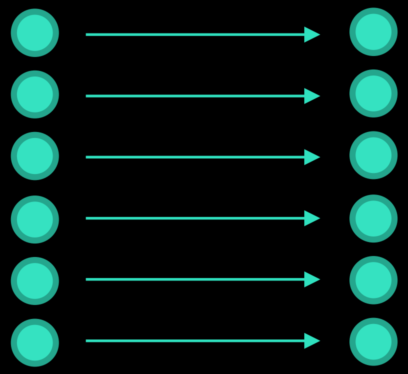


Discussion

- Self-augmentation vs dual learning

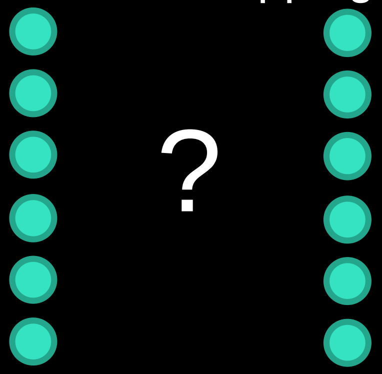
Traditional training

Rich data in both input and output
With known mapping



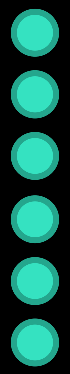
Dual learning

Rich data in both input and output
Learn the mapping



Self-augmentation

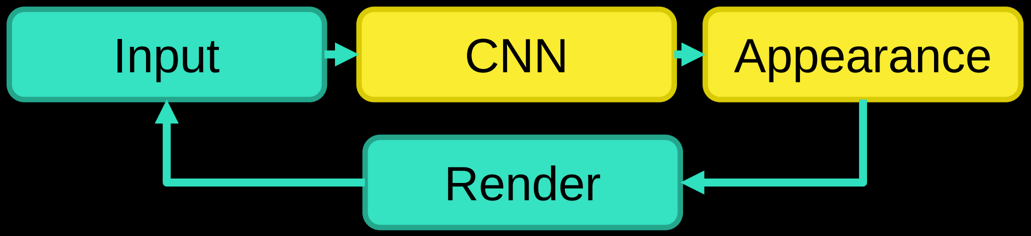
Rich data only for input



Discussion

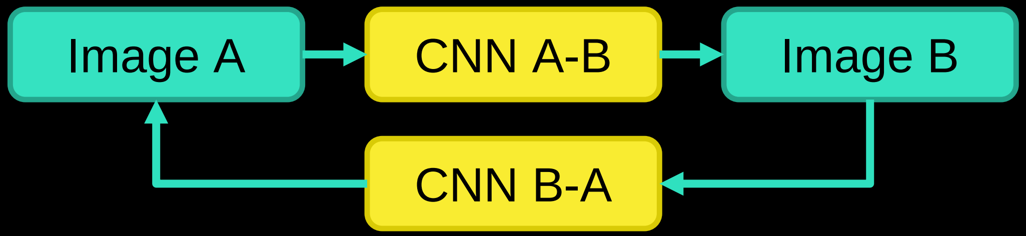
Self-augmented training

- Unlabeled input
- Known inverse mapping



Dual learning

- Unlabeled on two sets
- Trained dual tasks

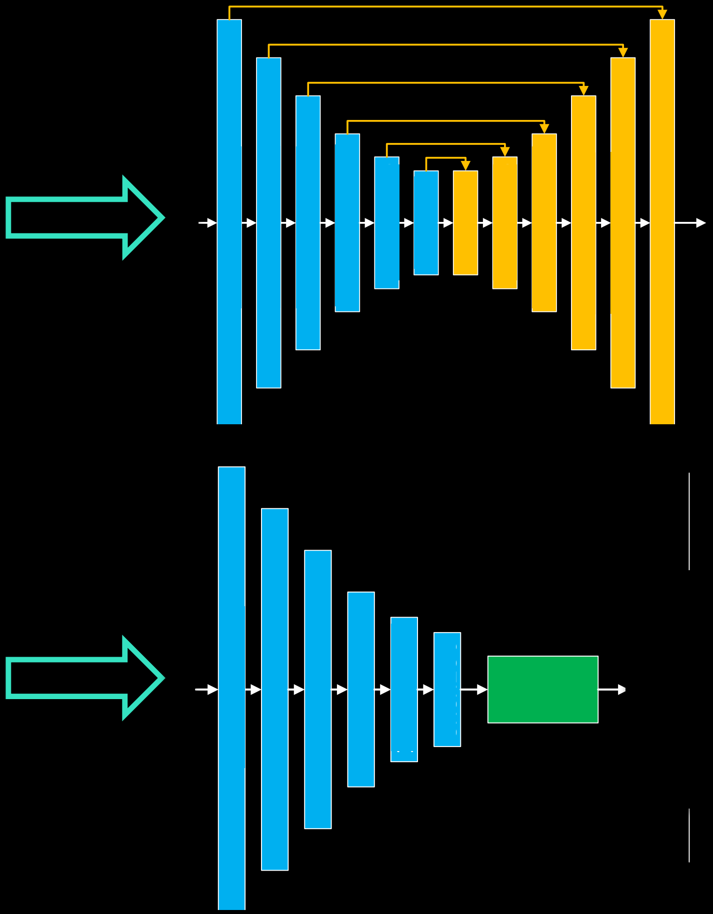


Motivation – Appearance Modeling

Tracer - by **Pyroshii** on DeviantArt / TommyGTeguh.com

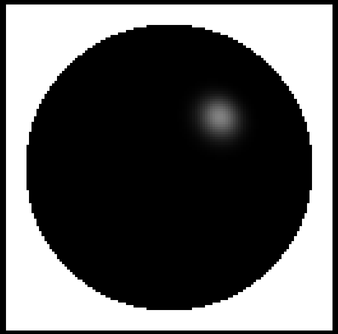
Modeling Appearance by CNN

Input Image



Albedo Map

Bump Map



Homogeneous Specular
(Ward Model)

BENEFIT of Self-Augmentation