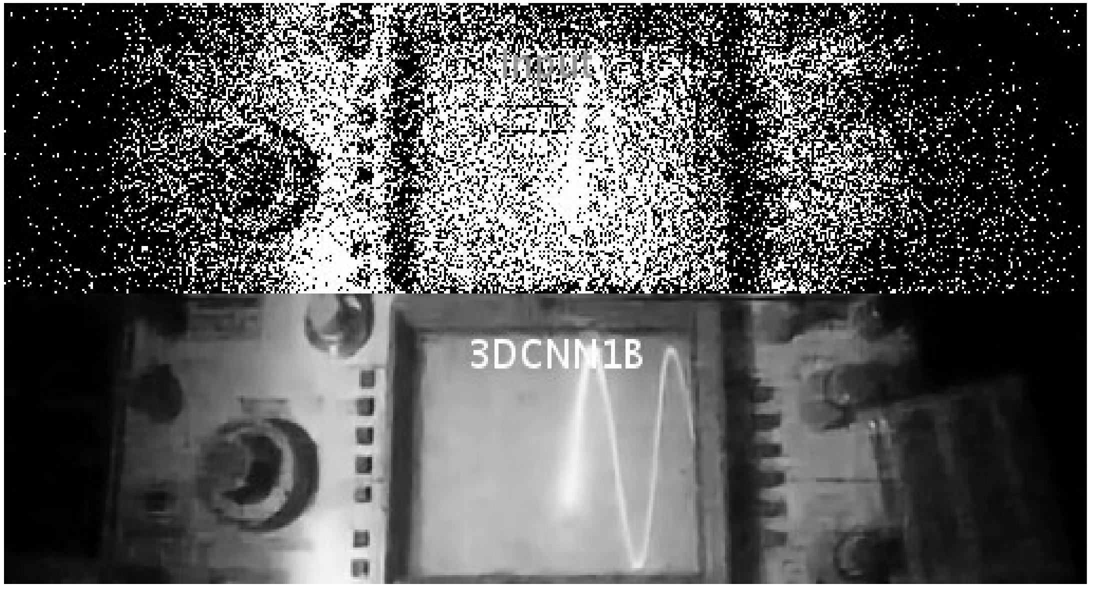
Deep Learning for Computational Photography

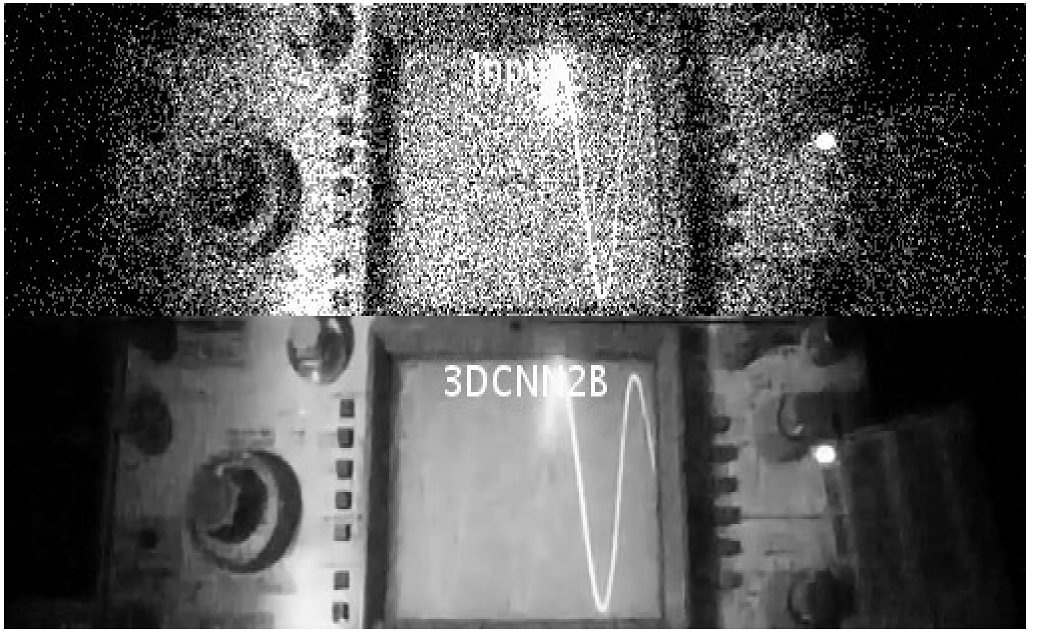
Paramanand Chandramouli
University of Siegen

High Speed Imaging using SPAD sensors

High Speed Imaging



High Speed Imaging



SPAD cameras

- Single photon avalanche diode (SPAD) sensors
- SwissSPAD- Programmable bit depth and high frame rates
- A/D conversion happens within each pixel (1-bit counter)
- Global gating of 5ns (for a 1-bit measurement) and readout (6.4 μs)
- Applications- Time-of-flight (ToF), Fluorescence Lifetime Imaging Microscopy (FLIM), Positron Emission Tomography (PET)

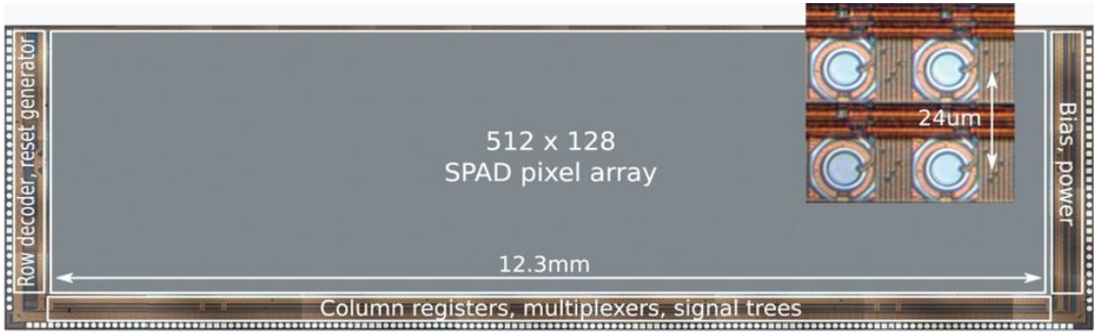


Image formation

- Photon count at a pixel is Poisson distributed

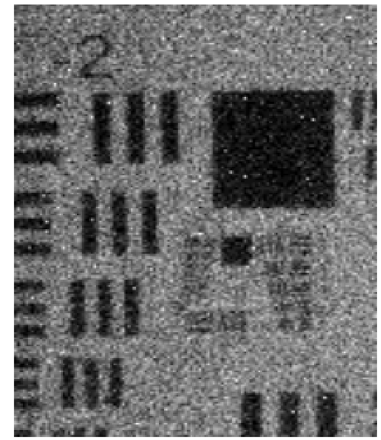
$$p_c(k) = \frac{\chi^k e^{-\chi}}{k!}$$

- χ denotes the expected value of counts per unit time
- $P(\text{count} > 0) = \ 1 - e^{-\chi}$
- Expected value of photon count is proportional to the scene radiance which gets non-linearly mapped at the sensors
- Image intensity $I_t(i,j)$ is a Bernoulli sample with a positive probability value of $1 - e^{-\chi}$

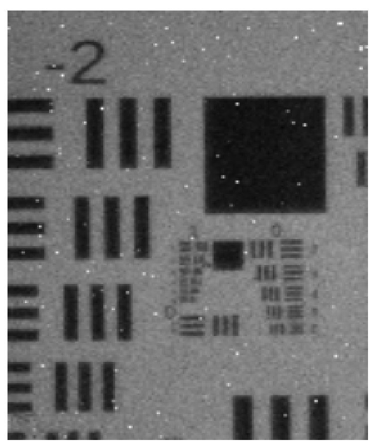
Image formation

1-bit

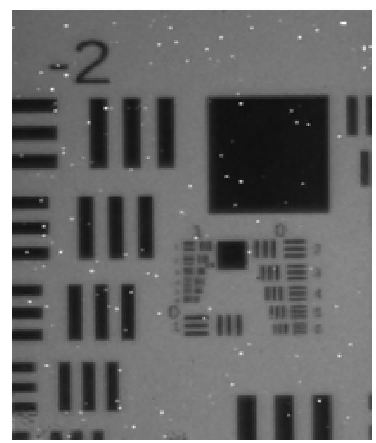
2-bit



4-bit



8-bit



16-bit

Problem formation

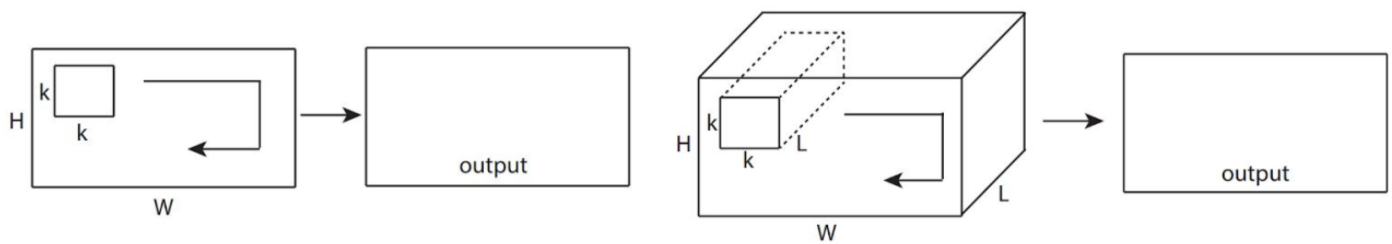
- SwissSPAD can capture 1-bit image sequences at the rate of 156 kfps
- A *b*- bit image sequence can be obtained by averaging N_b frames
- This results in a reduction of frame rate by a factor of N_b
- Objective:
  - Input- low bit depth high speed image sequence
  - Output- high bit depth estimate

Proposed Scheme

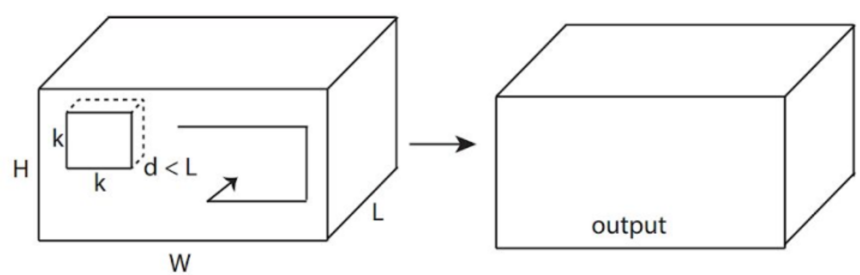
- Previous video denoising methods use hand-crafted approaches to combine information present in videos
- In an extreme scenario, hand-crafted approaches do not work
- We use a 3D CNN which effectively combines spatial and temporal information

Proposed Scheme

- 2D convolutional architecture- temporal information is not preserved

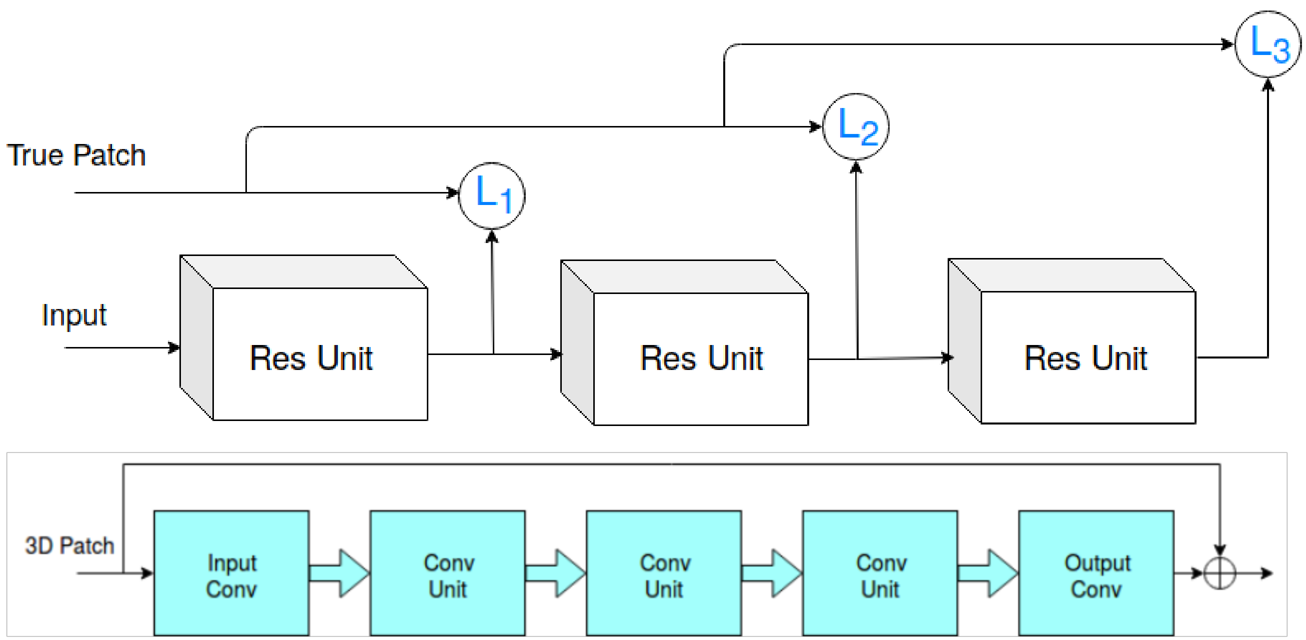


- 3D convolutional architecture preserves temporal information



Proposed Scheme

- Data sets for training: High speed videos to mimic temporal coherence
- ResNet-type architecture inspired by 2D image reconstruction
- Higher depth leads to larger receptive field
- Possible other architectures to optimally combine



High Speed Imaging

- Synthetic Results 1-bit

High Speed Imaging

- Synthetic Results 1-bit

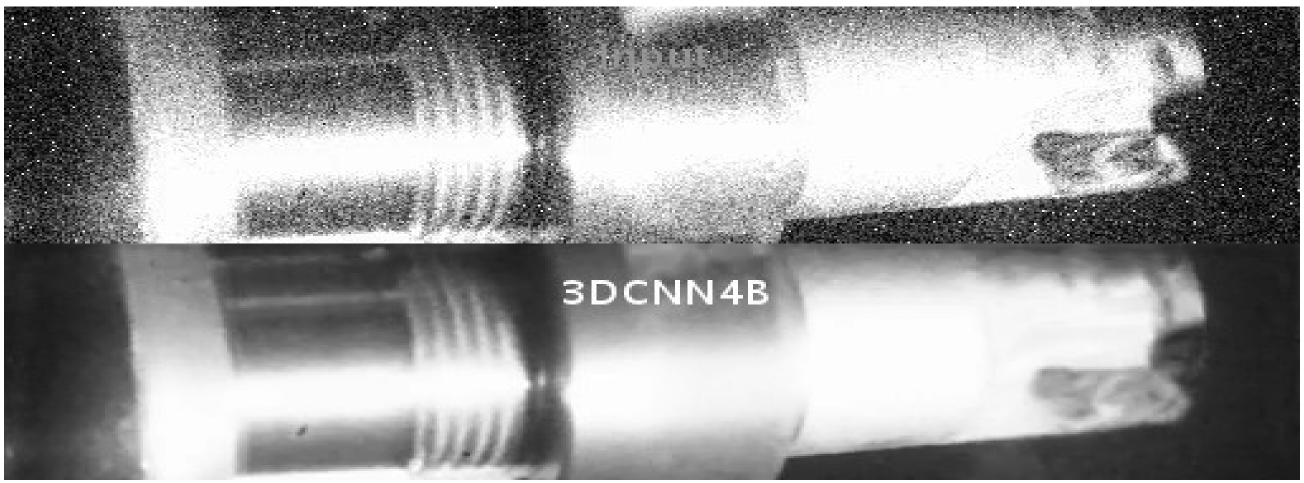
High Speed Imaging

- Synthetic Results 1-bit

High Speed Imaging

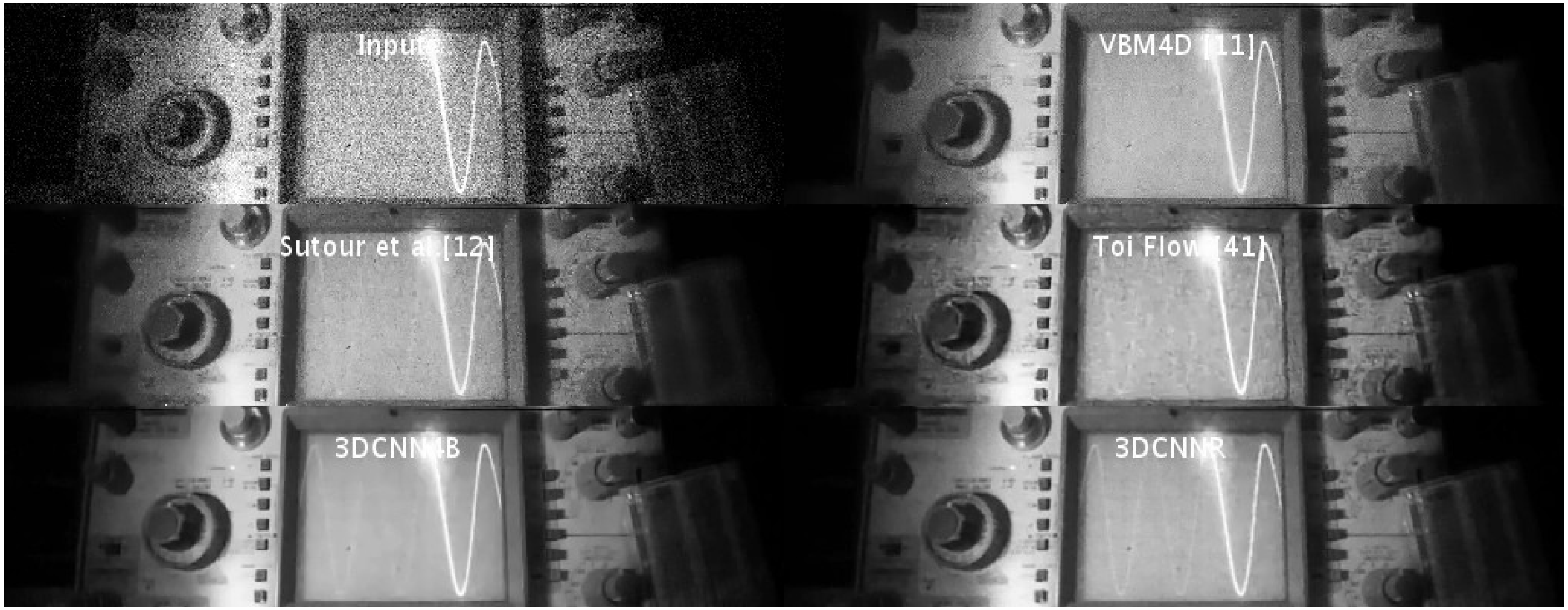
High Speed Imaging

•



High Speed Imaging

- 4-bit real



High Speed Imaging

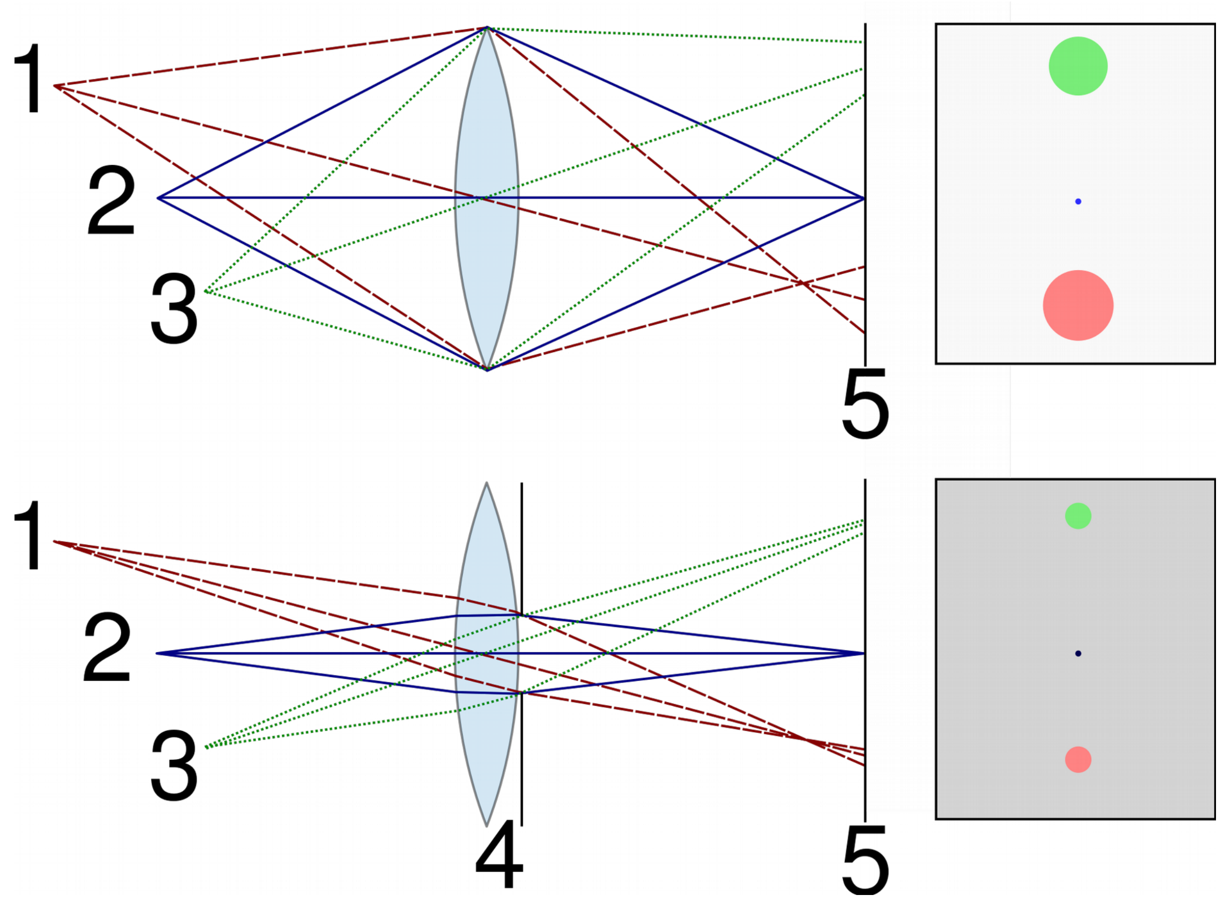
•

High Speed Imaging

- Video Denoising methods work for only 3-bit and 4-bit scenarios
- Even there, the proposed 3DCNN-based method is better
- For user, it is important to get a feedback as to which region is reliable
- Typically static regions are more reliable
- Can be determined by simple statistics like variance of residual
- Project: Re-implement in PyTorch

Light Field Imaging

Optical Defocus



Changing Focal Plane

Changing Focal Plane

Changing Focal Plane

3D Reconstruction

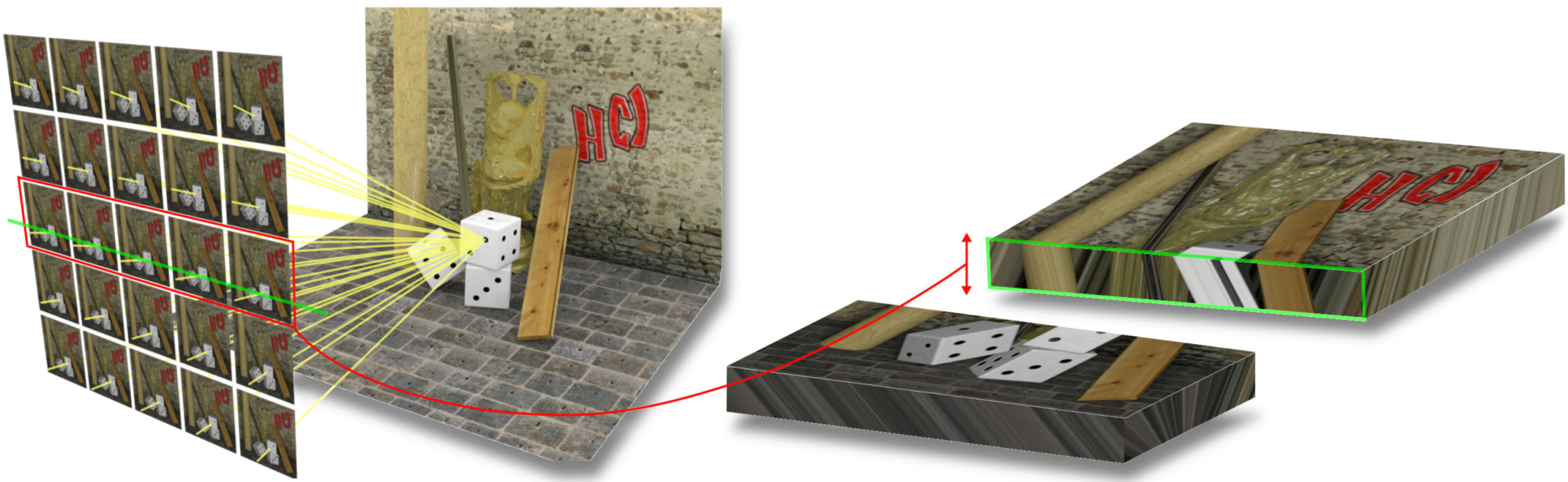
Plenoptic Cameras

Lytro Illum

Raytrix

Light Field

- Light field (LF) – function that describes the light intensity in every direction at every point in space



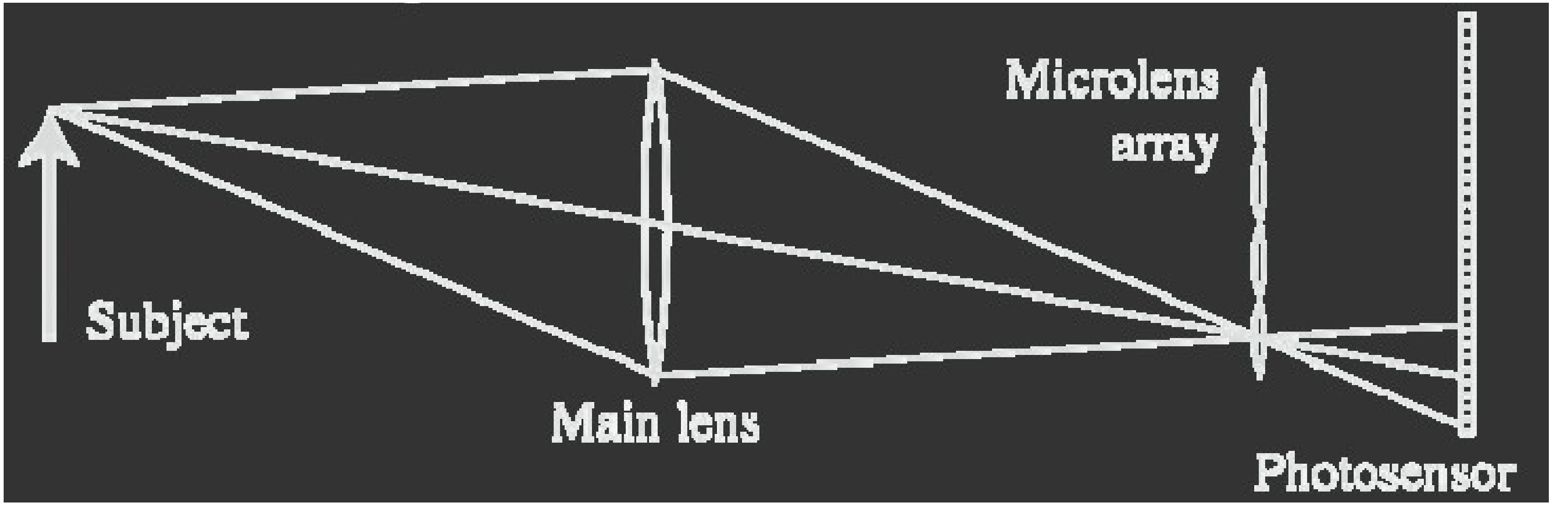
- Two plane parameterization- *L(u,v,s,t)*
- (s,*t*) pixel location (*u*,*v*) position of camera
- EPI representation: slope indicates disparity

Light Field

- Multi-view stereo with a small baseline
- Useful in refocusing, microscopy, VR, material properties

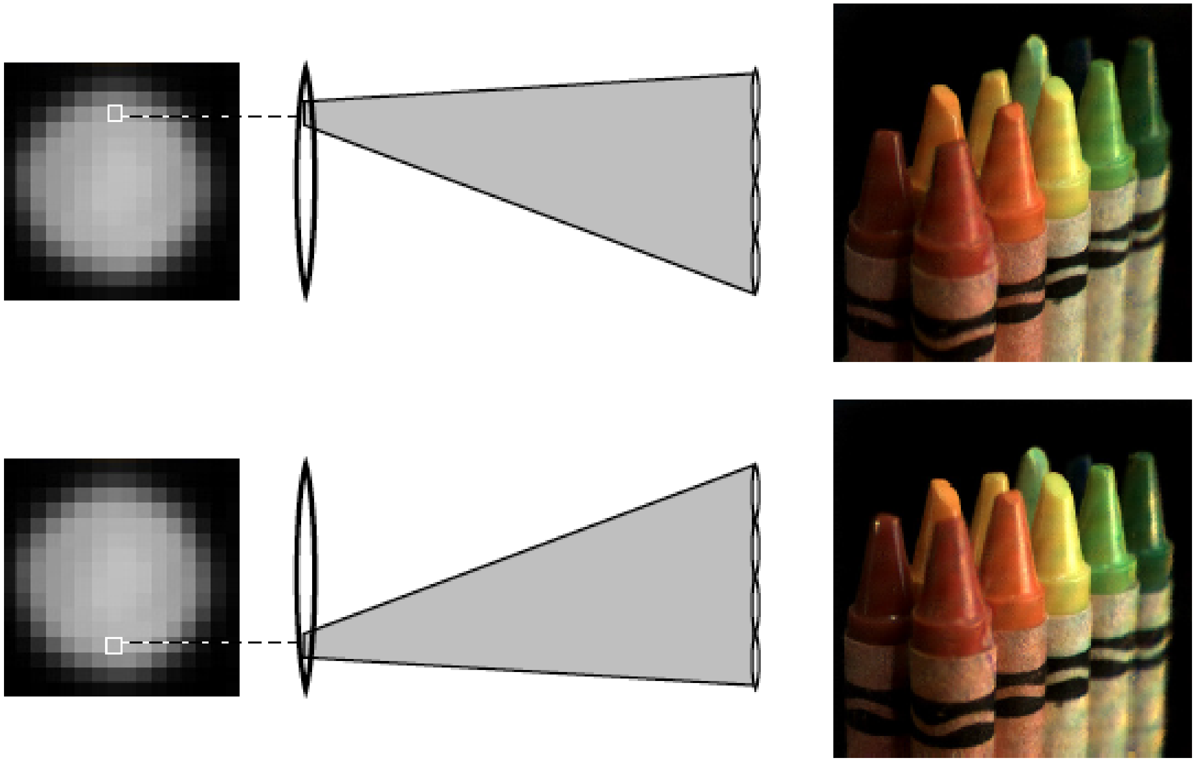
Lenslet-based LF camera

- LF from a snapshot of a single camera
- Plenoptic camera- microlens array between main lens and sensor plane



Lenslet-based LF camera

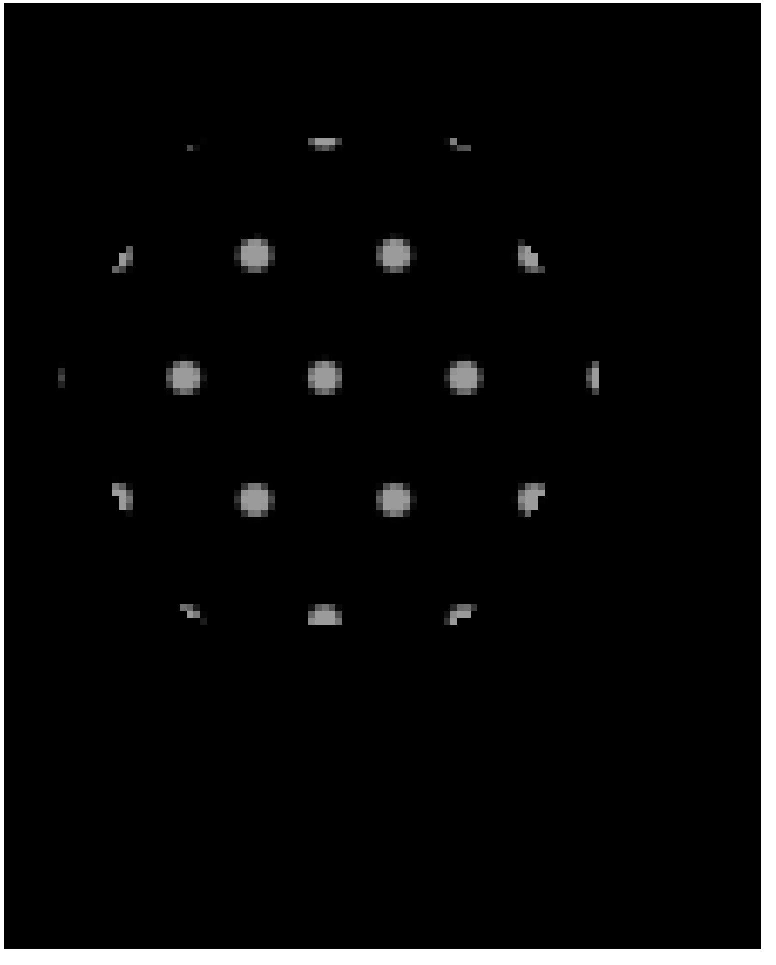
- These sub-aperture images can be used for refocusing and depth etimation
- Note the spatial resolution loss
- Baseline corresponds to aperture size



Point Spread Function (PSF)-based Modeling for Lambertian scenes

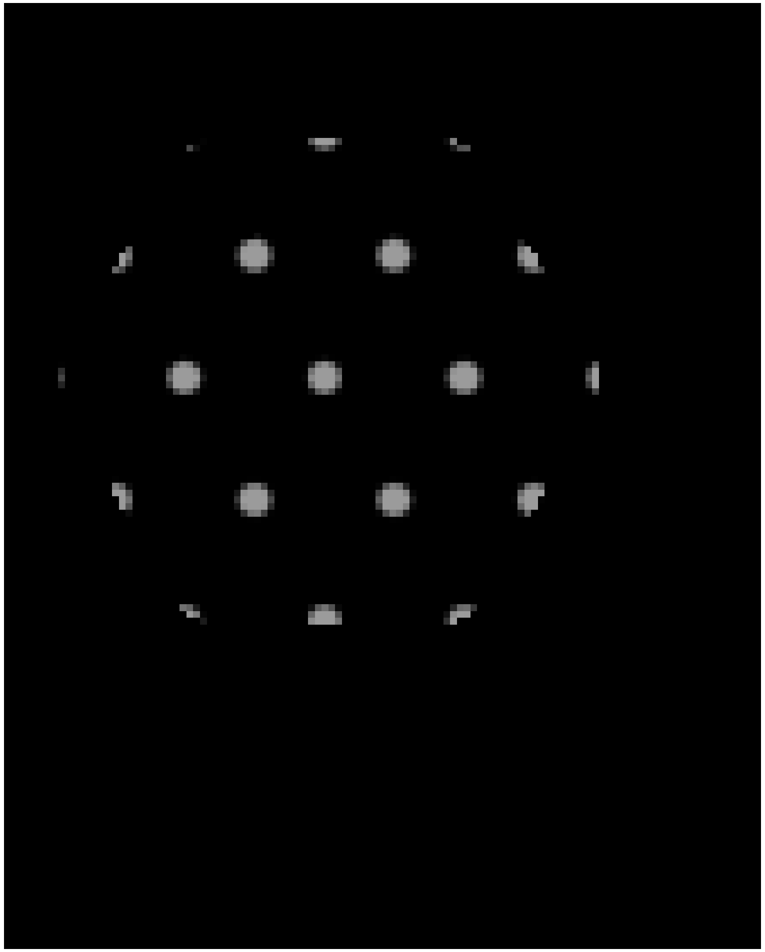
Point Spread Function

- Images obtained from a point light source



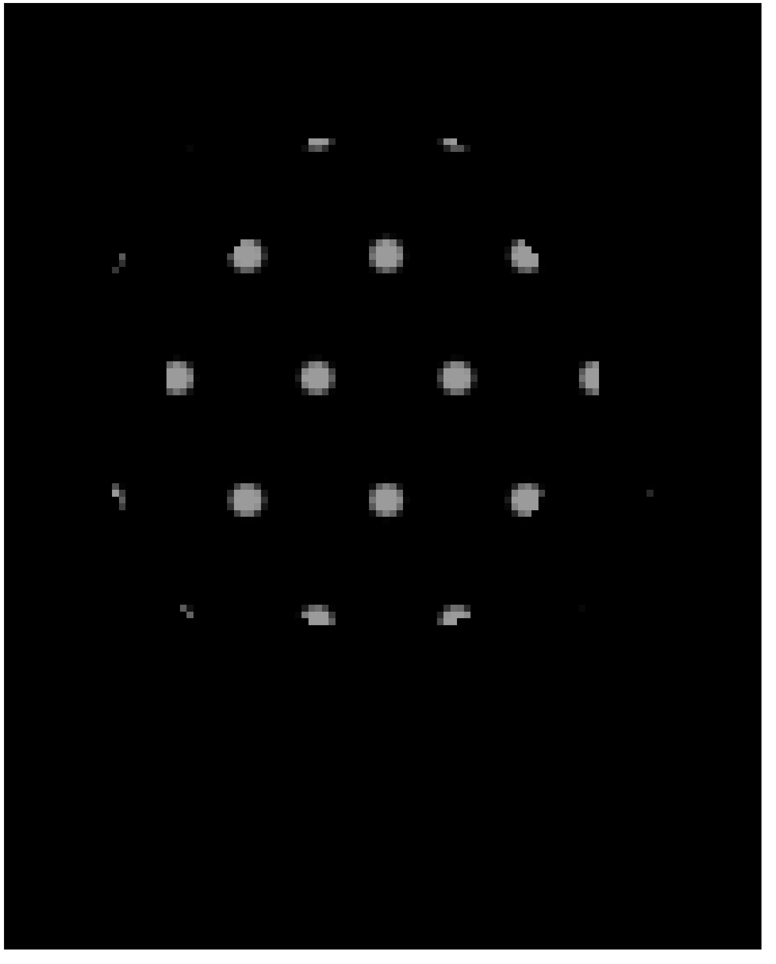
Point Spread Function

- Images obtained from a point light source



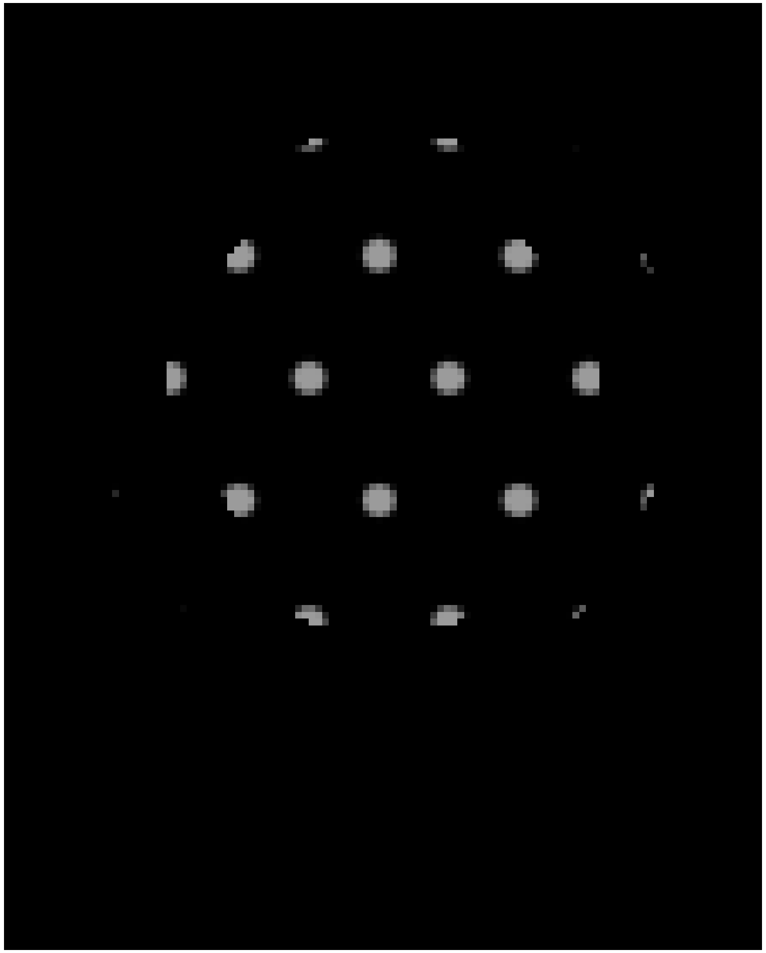
Point Spread Function

- Images obtained from a point light source



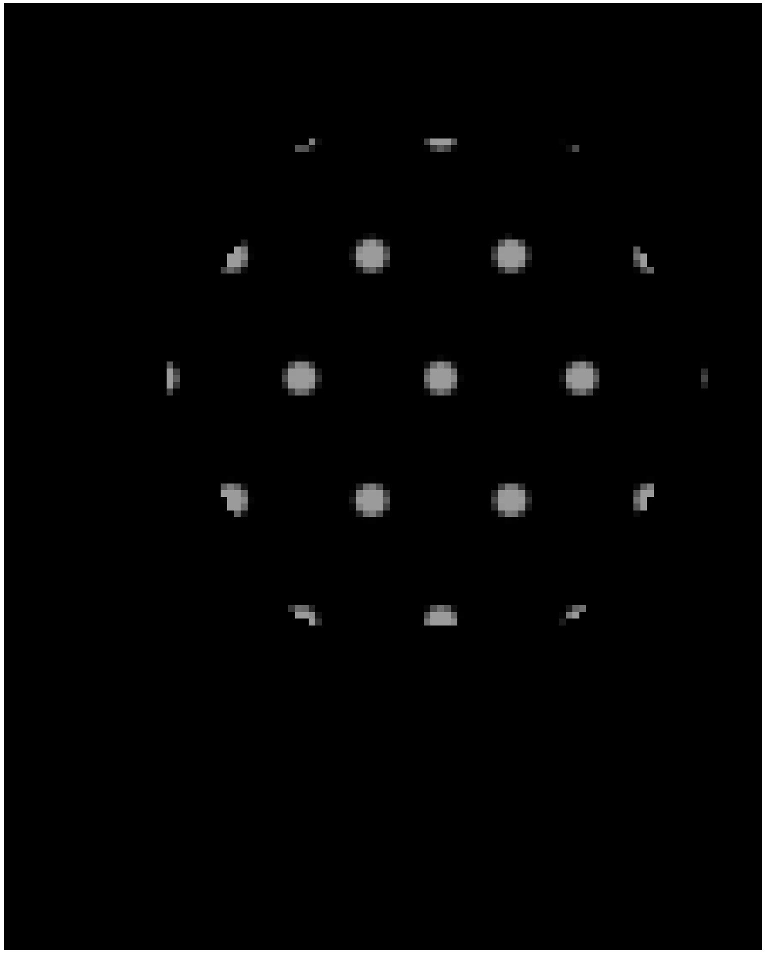
Point Spread Function

- Images obtained from a point light source



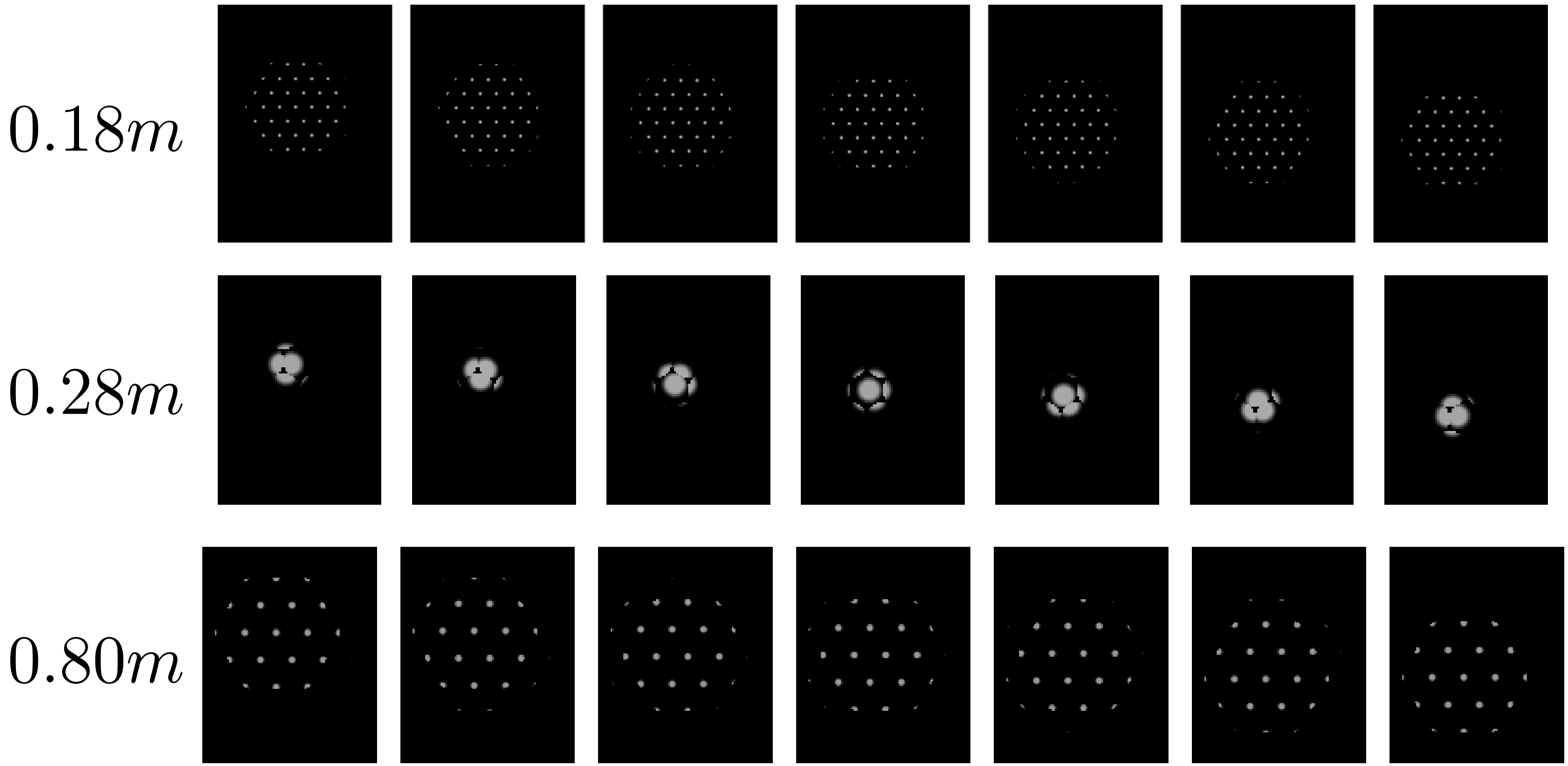
Point Spread Function

- Images obtained from a point light source



LF Image Formation

- Point spread function (PSF) varies with depth



PSFs: Images obtained by vertical shifting of point light sources

Layer Separation

Ill-posed

- One equation with two unknowns

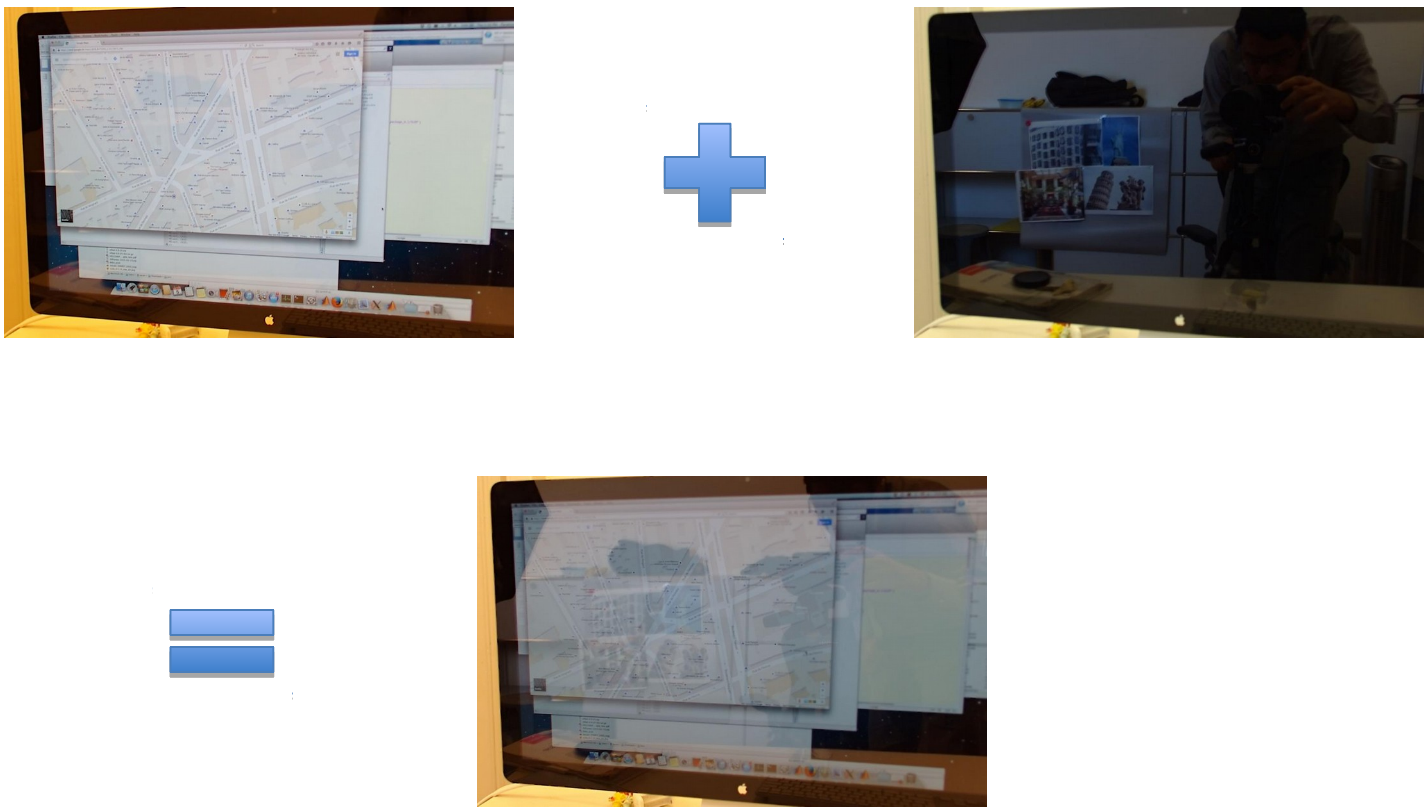
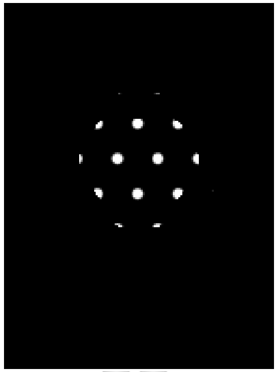
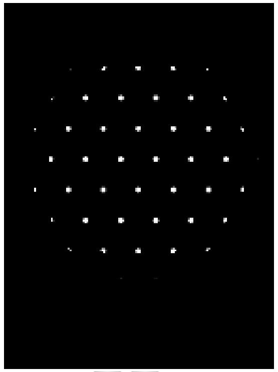


Image Formation

$$l = H_{d_t} f_{d_t} + H_{d_r} f_{d_r} = H f$$



H_{d_t}



H_{d_r}

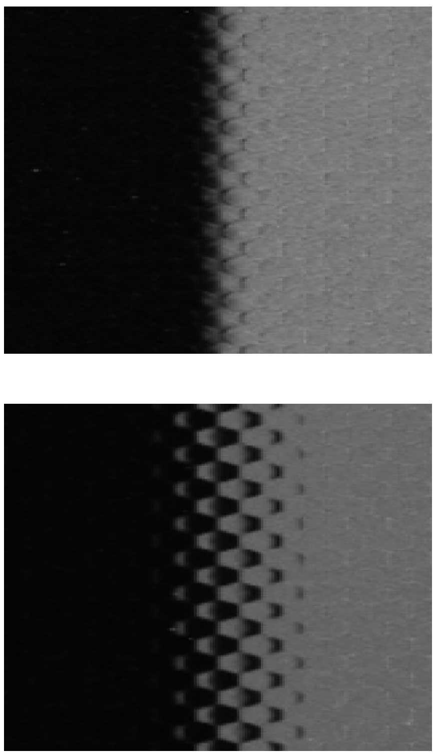
Unknowns: d_t, d_r, f_{d_t} and f_{d_r}

Two-Step Approach

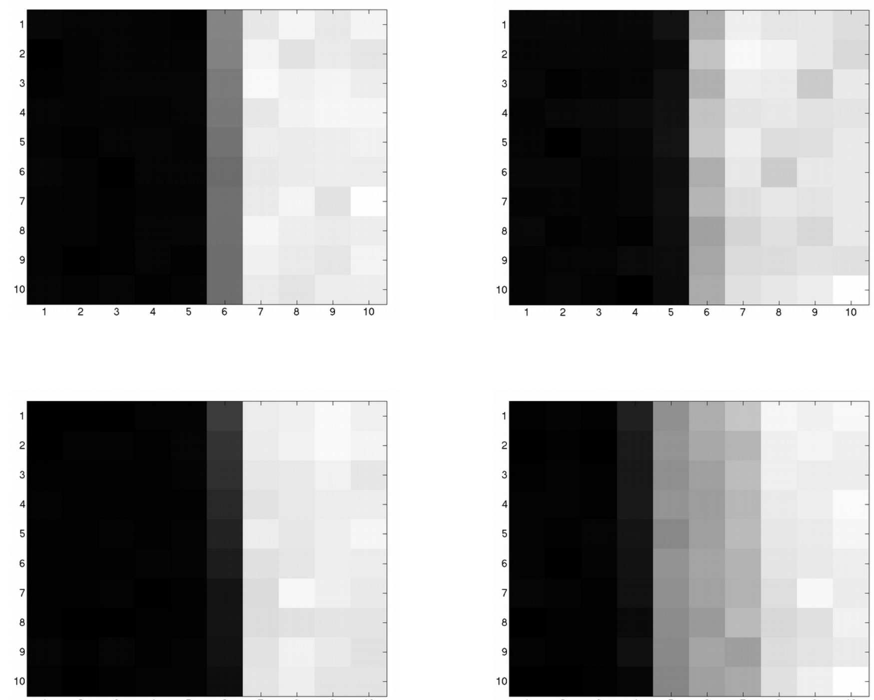
- LF enables depth estimation without the knowledge of texture

Depth Estimation

- Input – views sampled from patches of raw LF image are used



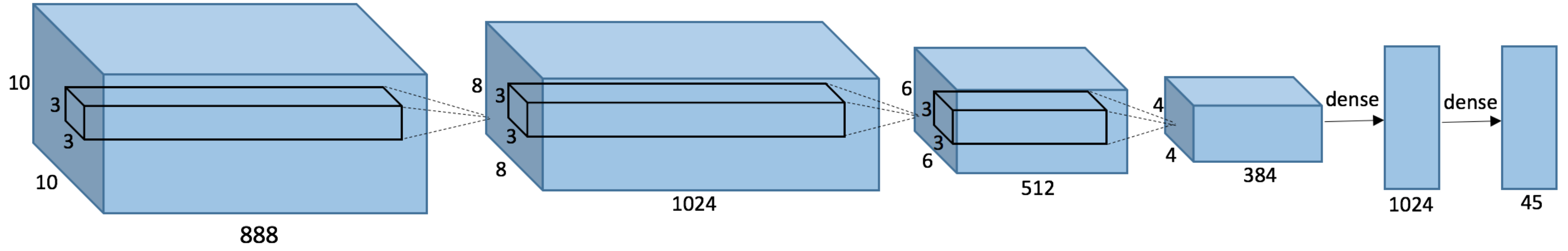
LF Patches



Views

Depth Estimation

- ConvNet-based classifier
- Label set consists of individual depths and combinations



ConvNet architecture

Depth Estimation - Lambertian

Raw image

Scene texture

Wang et al [ICCV 2015]

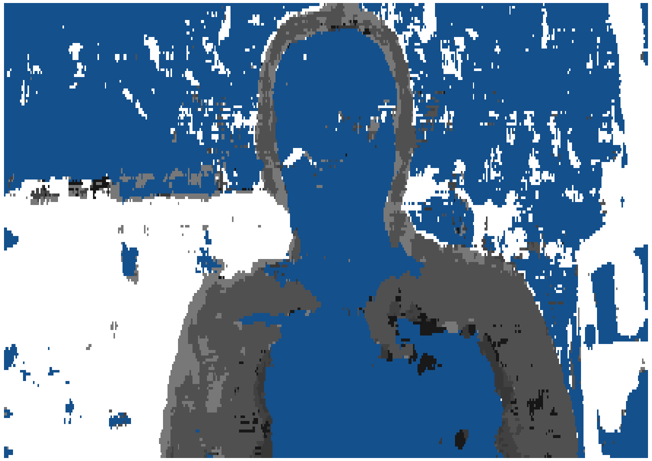
Proposed ConvNet-based

Depth Estimation - Reflective

Lytro rendering

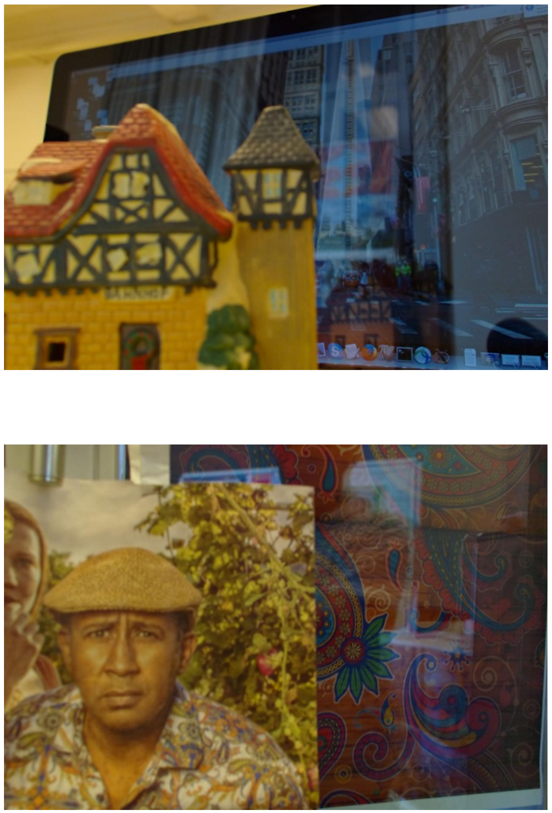
Wang et al [ICCV 2015]

Layer 1 depth



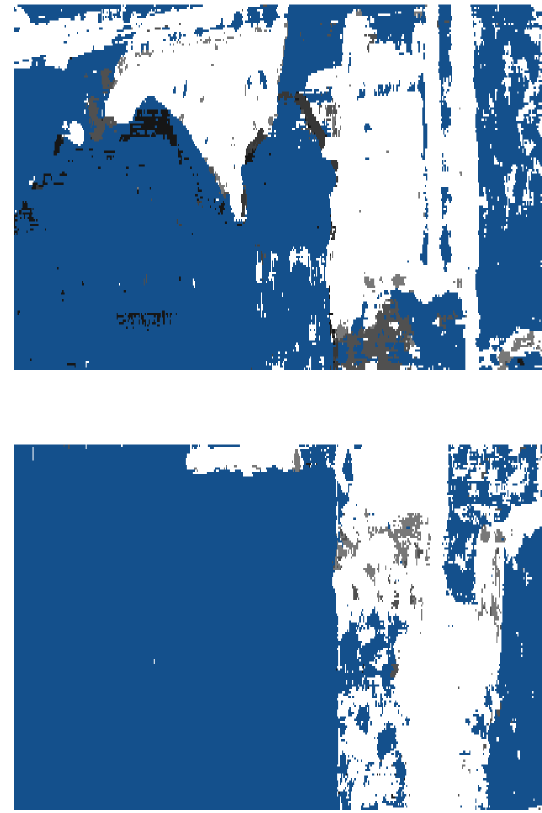
Layer 2 depth

Depth Estimation - Reflective



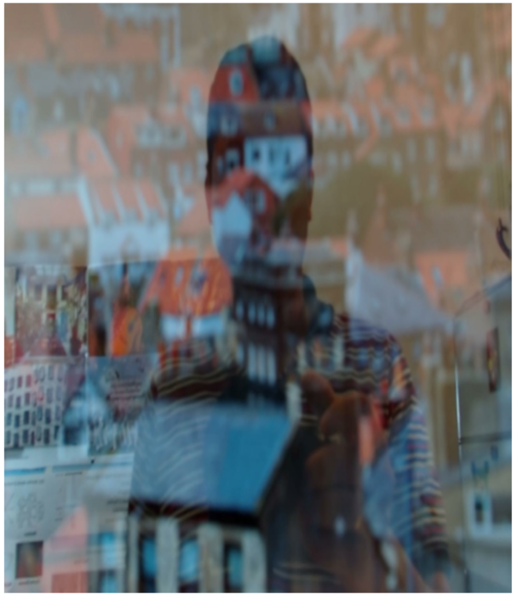
Lytro rendering

Layer 1 depth



Layer 2 depth

Layer Separation Results



Lytro rendering

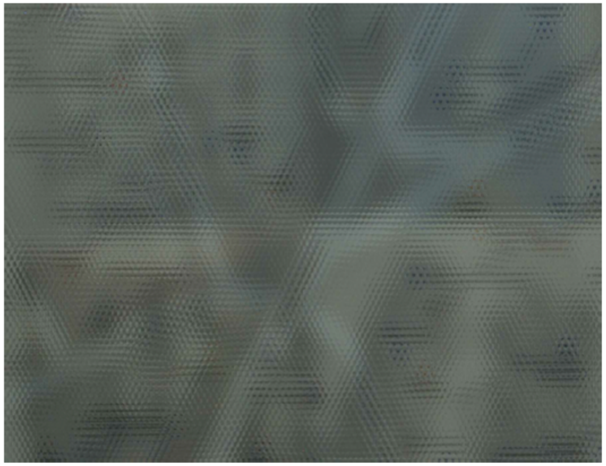
Recovered Layer1

Recovered Layer2

Layer1 depth

Layer2 depth

Layer Separation Results

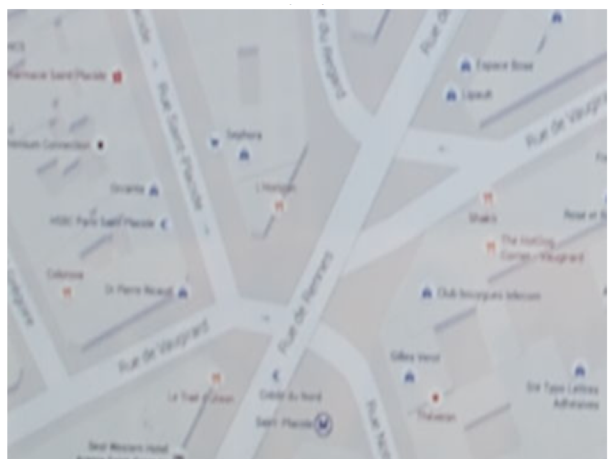


Raw Image

Lytro rendering

Recovered

Recovered



“True”

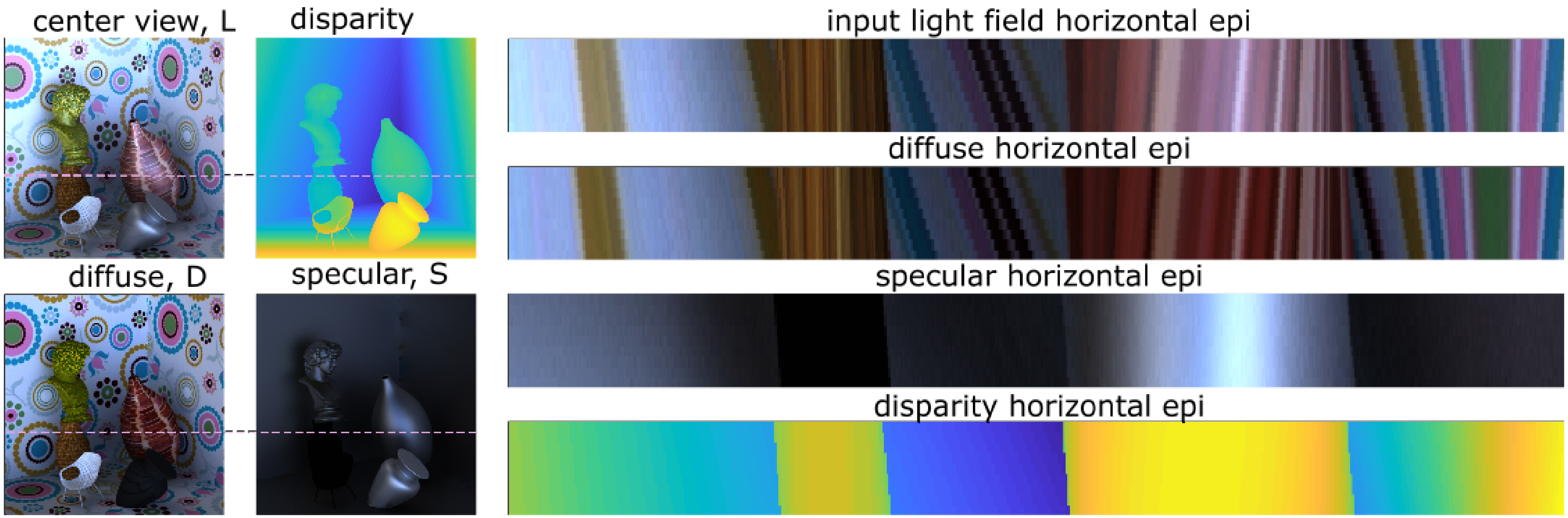
“True”

Light Field Intrinsics With a Deep Encoder-Decoder Network [Alperovich et al. CVPR 2018]

Autoencoder Model for LF

- Previously discussed depth estimation and many other LF depth estimation schemes need lot of labeled training data
- In this paper, a combination of both unsupervised and supervised training is used to reduce the required number of labeled data
- Address the tasks of disparity estimation as well as reflection separation

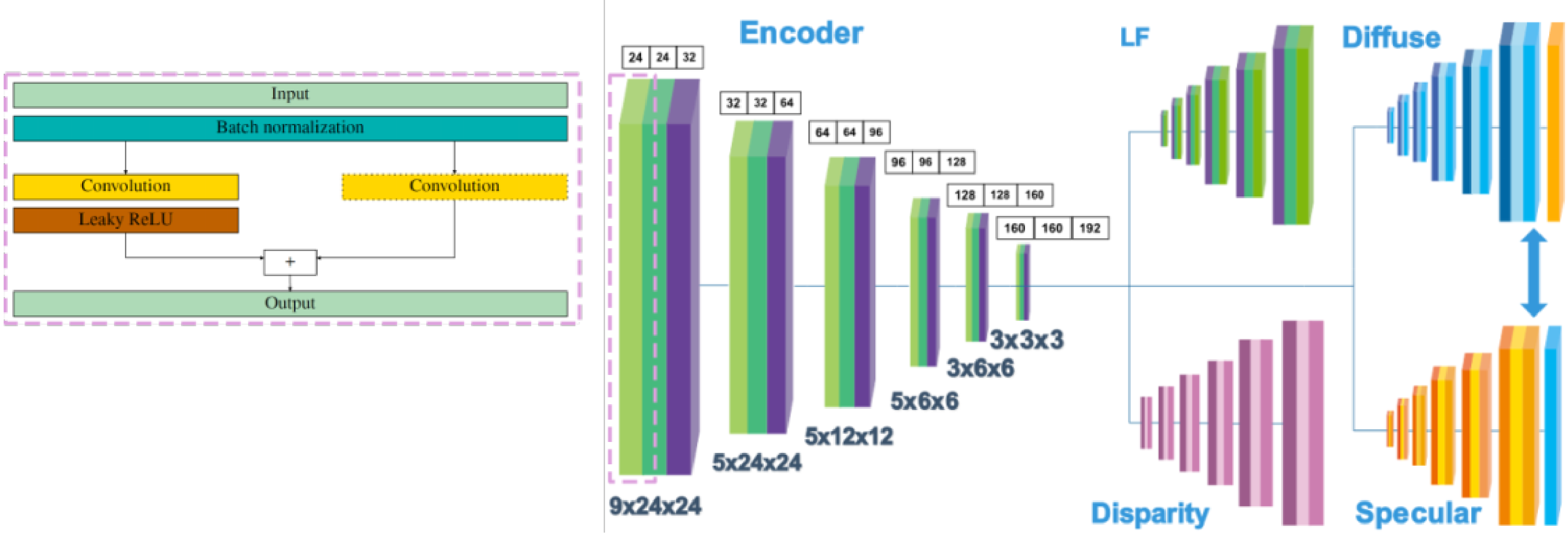
Dichromatic Reflection Model



$$L = D + S \quad (1)$$

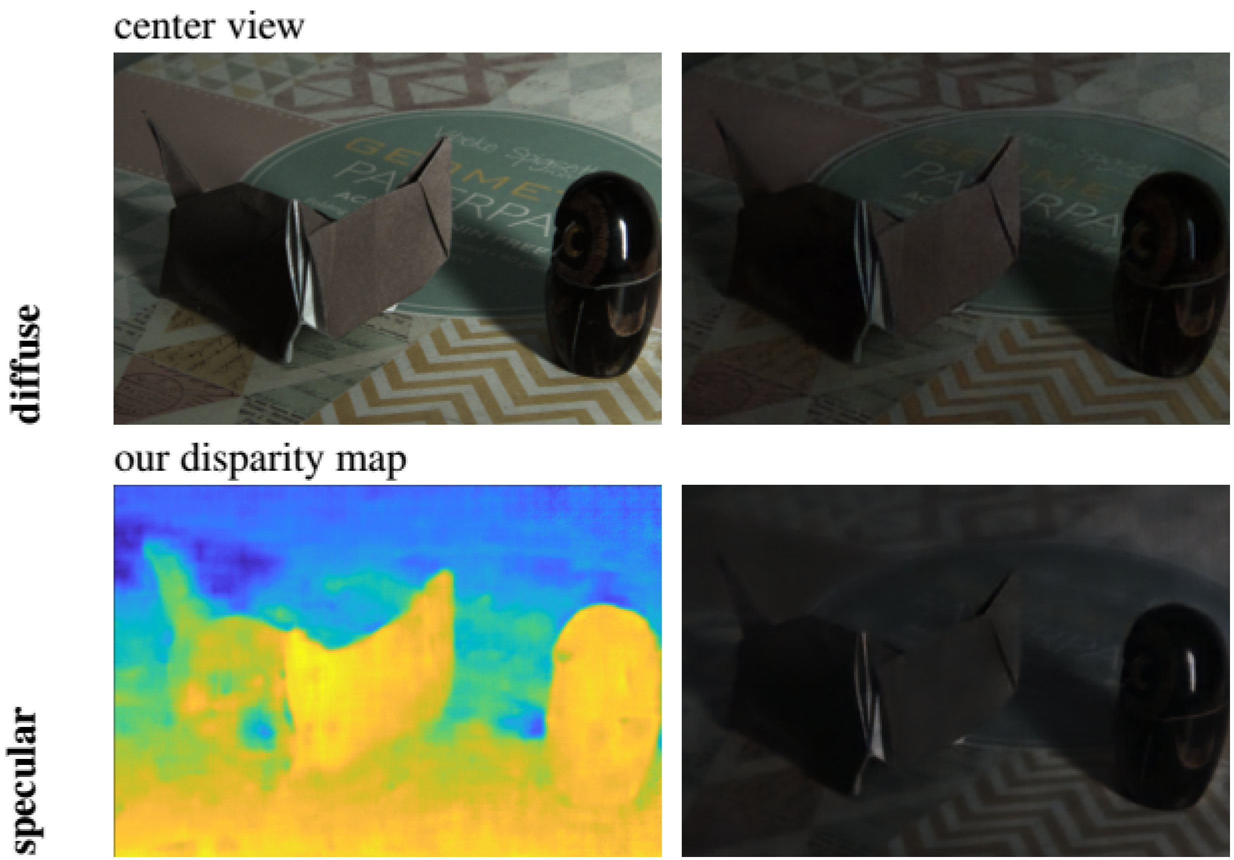
Network Architecture

- Multiple decoder paths trained depending on the availability of the data



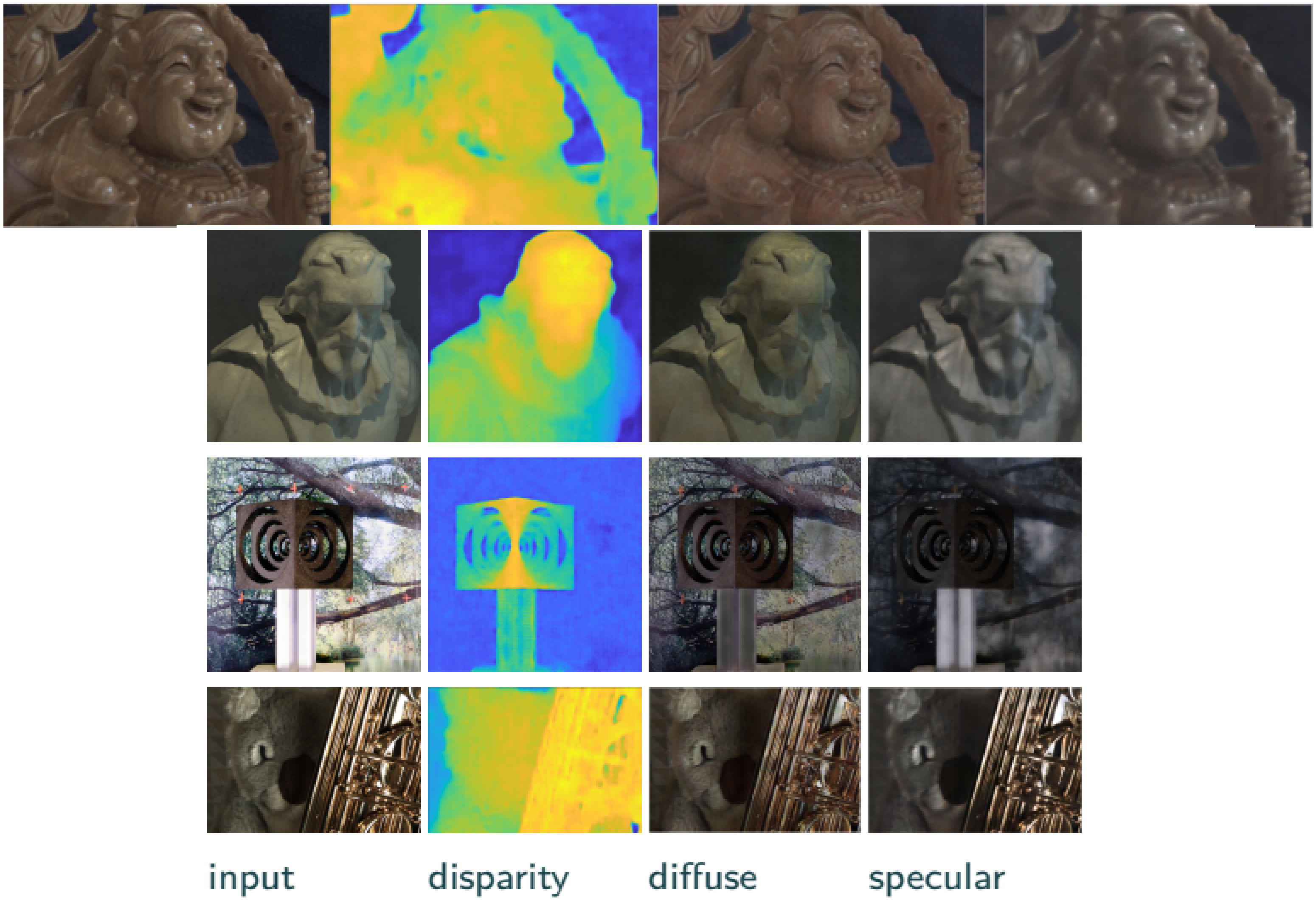
[Alperovich et al. CVPR 2018]

Results



[Alperovich et al. CVPR 2018]

Results



Thank You