TP.P4.5

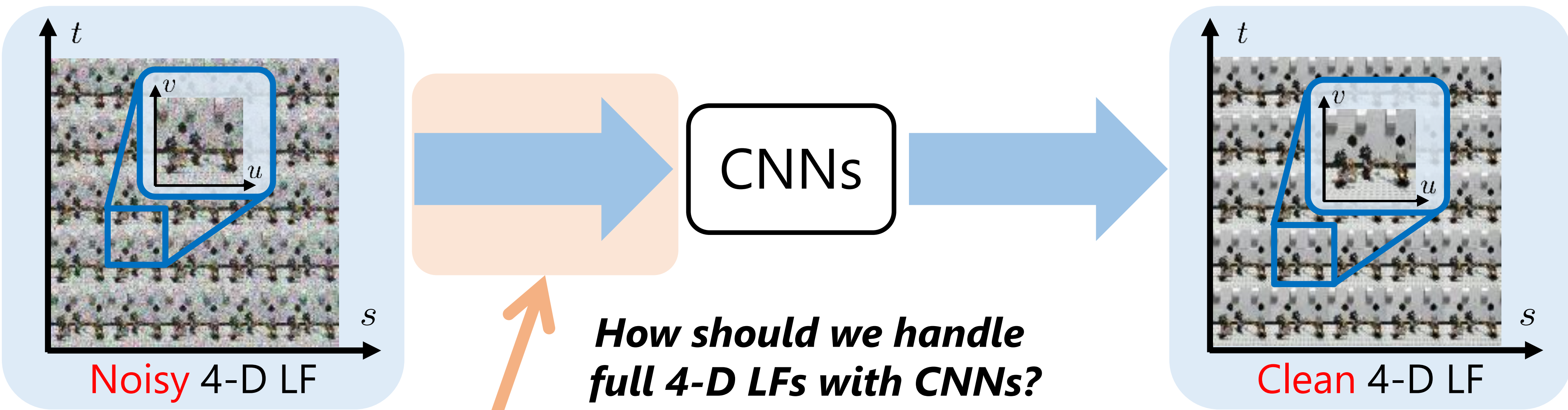
# How Should We Handle 4D Light Fields with CNNs?

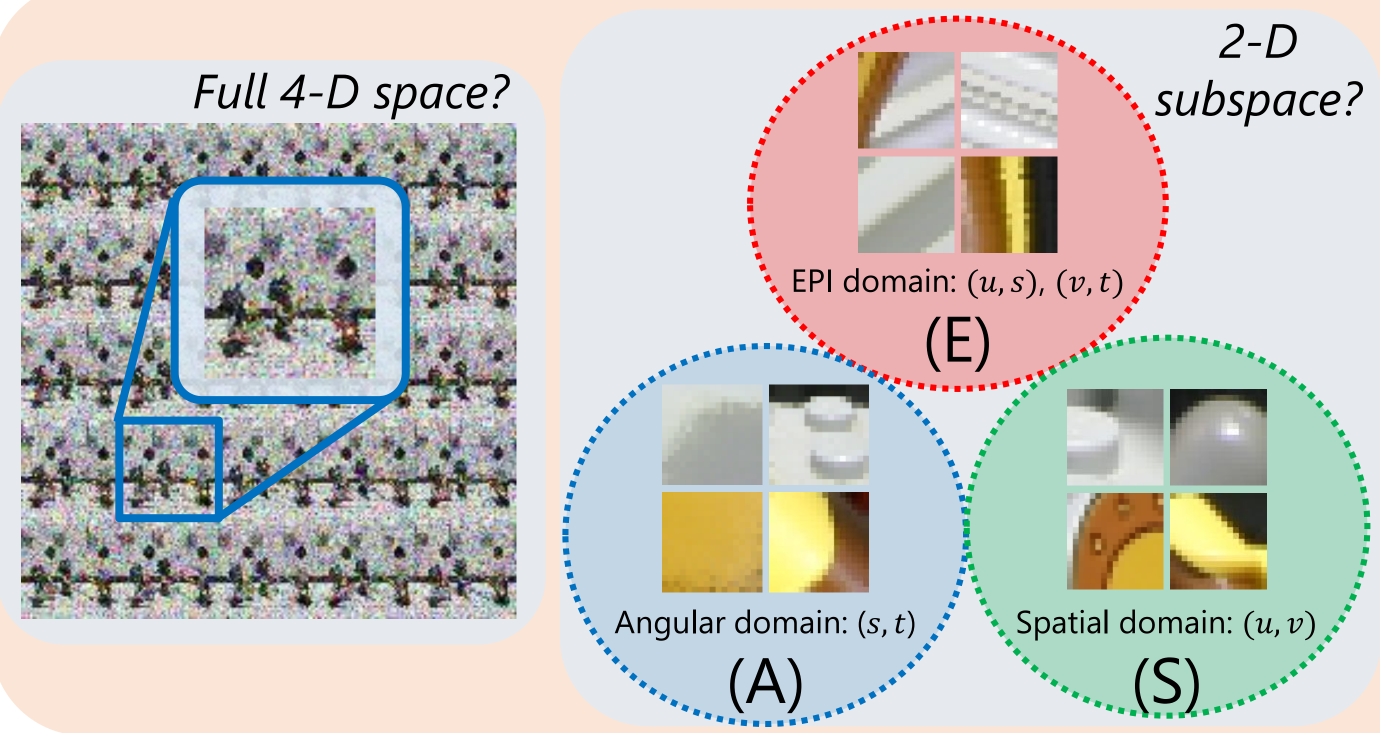
Shu Fujita, Keita Takahashi, Toshiaki Fujii     Graduate School of Engineering, Nagoya University, Japan

Fujii Laboratory

## Introduction

### Denoising with CNNs for 4-D Light Field Signals



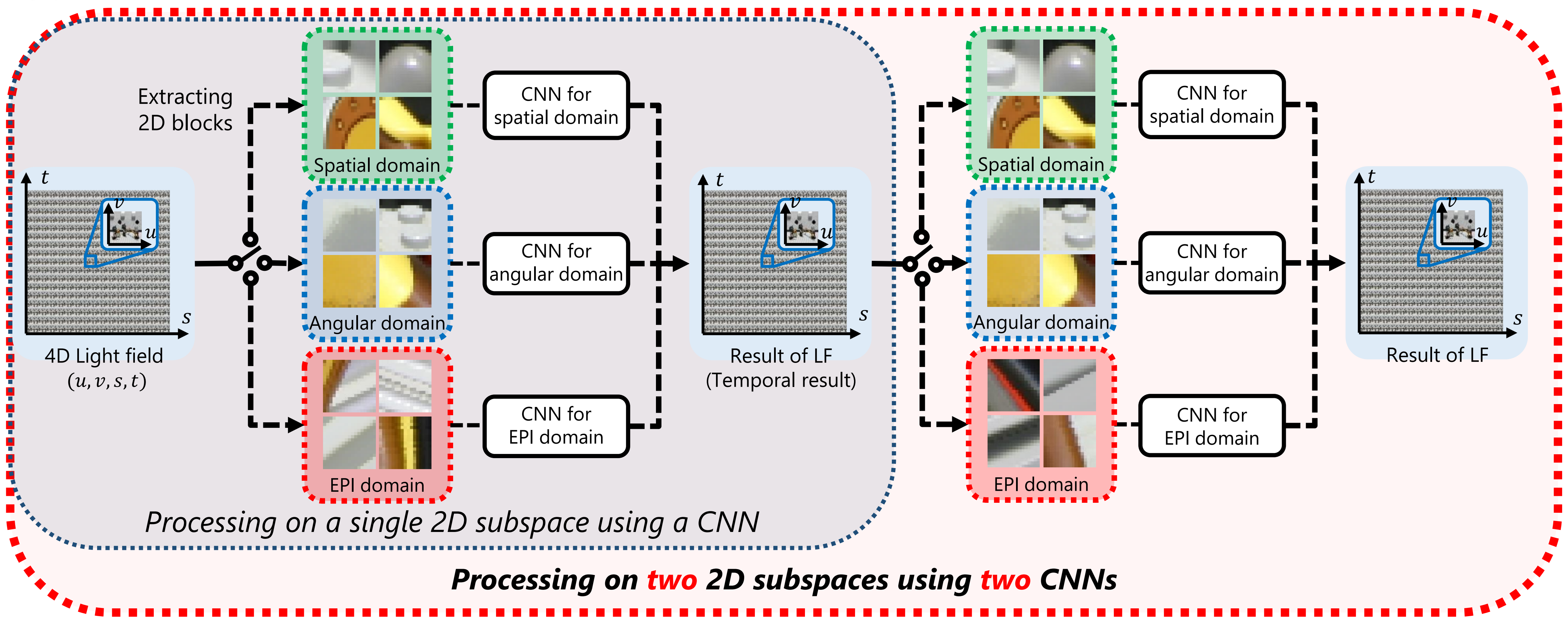

### Purpose

This question has not been thoroughly discussed.

We investigated how to handle high dimensional LFs, and proposed a pipeline for efficiently handling LFs.

## Cascading 2-D CNNs for 4-D Light Field

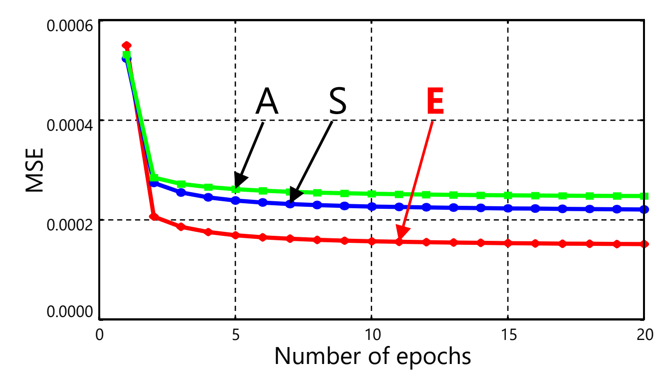
**Keypoint:** Cascading several CNNs, each of which works only on 2-D subspace of the full 4-D LFs



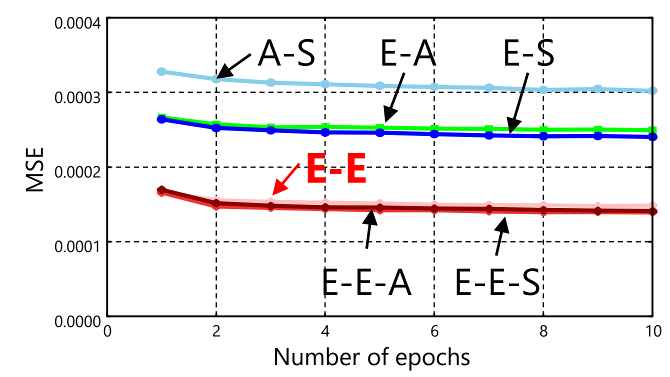

## Experiments

### Training Details

- CNN architecture: the same as VDSR [1], but the number of layers is different.
- The total number of layers for the entire process pipeline: 30
- Training dataset: gray-scale versions of the Stanford Light Field Archive
- Training samples: 10000 blocks (the block size is 17 × 17 × 17 × 17)
- The size of mini-batch: 200 (for a single CNN) and 10 (for cascaded CNNs)
- The number of epochs: 20 (for a single CNN) and 10 (for cascaded CNNs)
- Training loss and optimizer: MSE loss and Adam optimizer




(a) Case of using a single CNN.




(b) Case of cascading CNNs.

Fig. 1: Training loss along the epoch.

Tab. 1: PSNR [dB] of the entire LFs after denoising. (Gaussian noise is added)

| | Tarot | | Buddha | | Mona | | Papillon | | StillLife | |
|---|---|---|---|---|---|---|---|---|---|---|
| Noise std. dev. | 0.10 | 0.15 | 0.10 | 0.15 | 0.10 | 0.15 | 0.10 | 0.15 | 0.10 | 0.15 |
| Noisy input | 20.01 | 16.48 | 20.00 | 16.48 | 20.00 | 16.48 | 20.00 | 16.48 | 20.00 | 16.48 |
| **E** | **31.92** | **28.8** | **33.29** | **30.69** | **33.50** | **30.96** | **34.51** | **31.72** | **28.08** | **25.50** |
| A | 28.21 | 25.71 | 32.44 | 30.11 | 33.26 | 30.78 | 34.07 | 31.50 | 25.97 | 23.39 |
| S | 28.21 | 25.77 | 31.47 | 29.52 | 31.96 | 29.99 | 33.16 | 31.07 | 25.96 | 24.01 |
| **E-E** | **35.47** | **32.93** | **37.60** | **35.31** | **37.99** | **36.13** | **38.89** | **37.09** | **31.41** | **28.58** |
| E-A | 32.99 | 30.53 | 35.65 | 33.47 | 36.25 | 34.06 | 37.32 | 35.40 | 28.82 | 26.12 |
| E-S | 33.11 | 30.75 | 35.41 | 33.37 | 35.25 | 33.56 | 36.99 | 35.25 | 29.03 | 26.47 |
| A-S | 30.64 | 28.22 | 36.13 | 34.40 | 36.73 | 35.23 | 37.91 | 36.51 | 26.73 | 24.674 |
| E-E-A | 35.48 | 33.24 | 37.69 | 35.60 | **38.07** | **36.43** | **39.13** | 37.39 | **31.60** | 28.88 |
| E-E-S | **35.63** | **33.34** | **37.94** | **35.80** | 37.97 | 36.37 | **39.13** | 37.60 | 31.56 | **29.27** |
| NLMF [2] | 26.06 | 23.27 | 27.96 | 24.86 | 28.07 | 24.90 | 28.36 | 25.06 | 24.83 | 22.52 |
| BM3D [3] | 28.53 | 26.13 | 32.33 | 30.33 | 32.80 | 31.05 | 34.49 | 32.51 | 27.18 | 24.93 |
| VBM4D [4] | **30.34** | **27.94** | **33.91** | **31.90** | **34.29** | **32.54** | **36.11** | **34.31** | **28.81** | **26.41** |

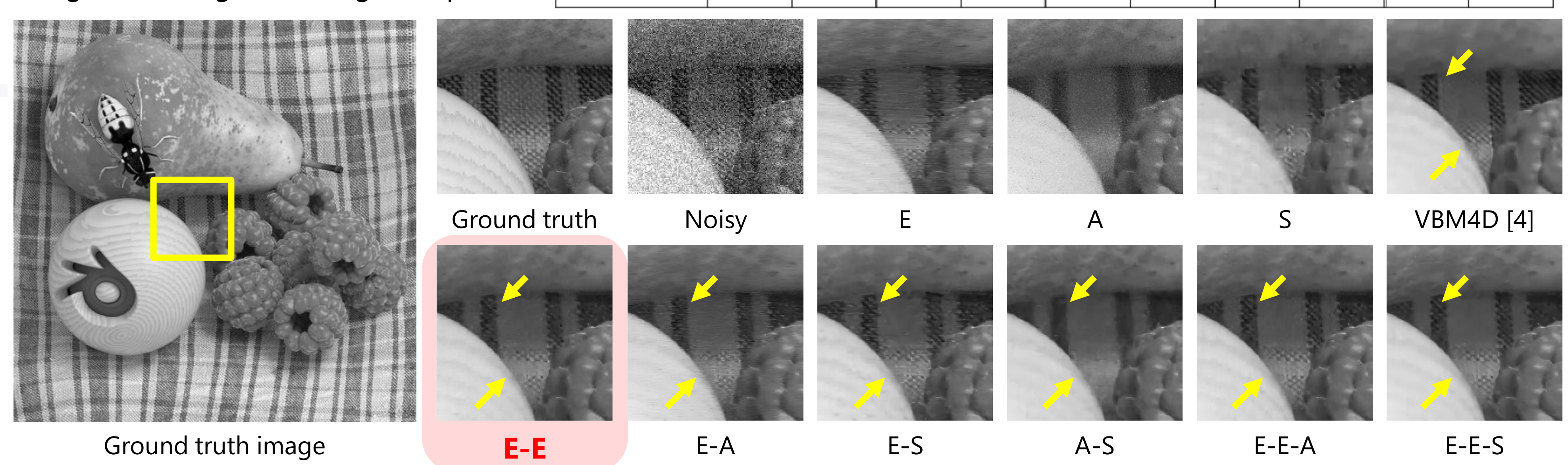



Fig. 2: Denoised images at center viewpoint. (We added Gaussian noise with standard deviation 0.10.)

## Summary

- The EPI domains are easier to learn than the other domains.
- The combination of two EPI domains is better than the others.

### Reference

[1] J. Kim *et al.*, "Accurate image super-resolution using very deep convolutional networks," in IEEE CVPR, 2016.
[2] A. Buades *et al.*, "A non-local algorithm for image denoising," in IEEE CVPR, 2005.
[3] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3-D transform-domain collaborative filtering," IEEE TIP, 2007.
[4] M. Maggioni *et al.*, "Video denoising, deblocking and enhancement through separable 4-D nonlocal spatiotemporal transforms," IEEE TIP, 2012.