

# Learning Non-Local Spatial-Angular Correlation for Light Field Image Super-Resolution

## 概要

- タスクはLight Field Image Super Resolution
- TransformerベースのEPITを提案

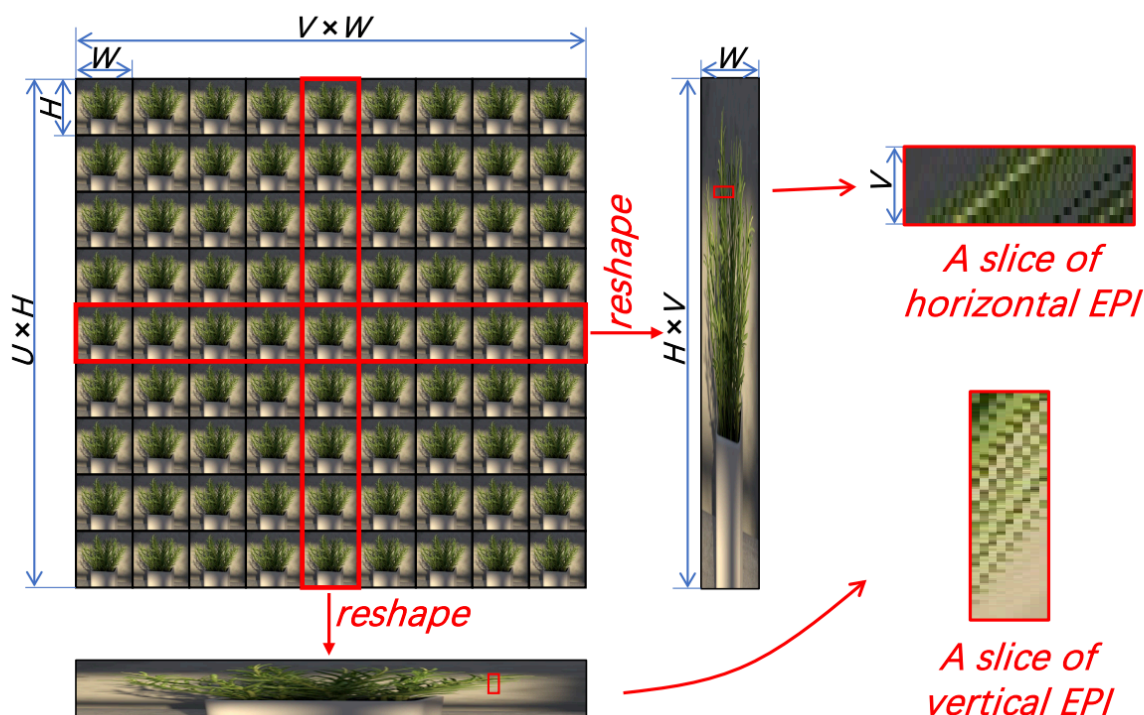
## 前提

### LF SR(Light Field Image Super Resolution)

- LF imageは4次元の関数として記述する

$$\mathcal{L}(u, v, h, w) \in \mathbb{R}^{U \times V \times H \times W}$$

- $U, V$ はangular dimensions(つまり、light fieldカメラの個数)
- $H, W$ はspatial dimensions



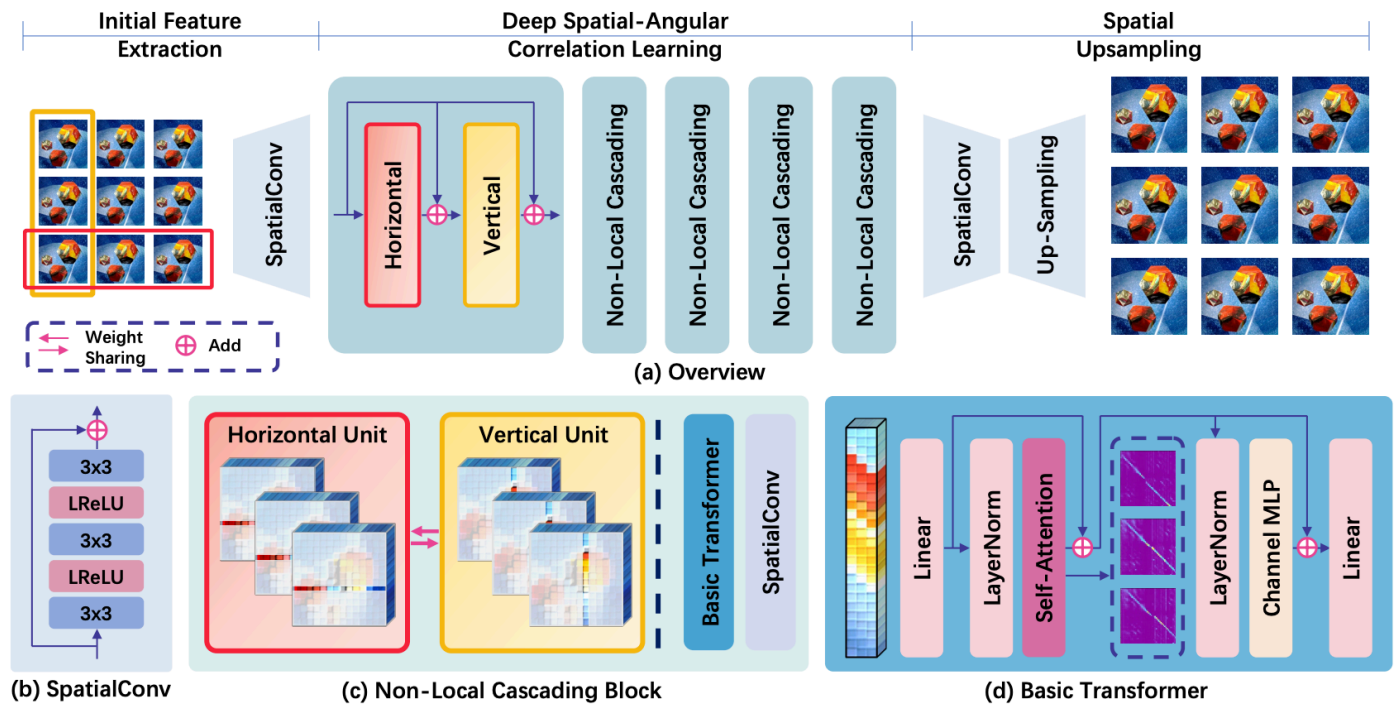
- よって、LR LFとHR LFはそれぞれ以下のように記述できる

$$\begin{aligned}\mathcal{L}_{LR} &\in \mathbb{R}^{U \times V \times H \times W} \\ \mathcal{L}_{HR} &\in \mathbb{R}^{U \times V \times \alpha H \times \alpha W}\end{aligned}\quad (1)$$

- LR SRではこれらを入力、出力とするタスク

## EPIT

- 大きく3層から構成される
  - initial feature extraction
  - deep spatial angular correlation learning
  - feature upsampling



## Initial Feature Extraction

- $3 \times 3$ のconvolutionとLeakyReLUから構成される
- 出力はchannel dimensionを $C$ とすると $F \in \mathbb{R}^{U \times V \times H \times W \times C}$

## Deep Spatial Angular Correlation Learning

- 5層のNon-Local Cascadingから構成される
- Non-Local Cascadingは2層のBasicTransformerとSpatialConvから構成される

- このとき2層のBasicTransformerはそれぞれHorizontal, Verticalと呼ばれる
- この2層のBasicTransformerであるHorizontalとVerticalはweightをshareしている
- 前層から得られた  $F \in \mathbb{R}^{U \times V \times H \times W \times C}$  をHorizontal(Vertical)に入力するときに  $F_{hor} \in \mathbb{R}^{UH \times V \times W \times C}$  ( $F_{ver} \in \mathbb{R}^{VW \times U \times H \times C}$ )にreshapeする
- reshapeしてtransformerに入れて得られたものを  $\hat{F}_{hor}$  ( $\hat{F}_{ver}$ )とする
- 次のSpatialConvに入れる前にこれらを  $UV \times H \times W \times C$ にreshapeする

## Feature Upsampling

- pixel shuffling operatinから構成される

## 損失関数

- lossはL1 loss