Implicit neural representation learning for hyperspectral image super-resolution

Arxiv. 2112

Haneol Lee

4th April, 2022

Contribution

- **We propose a novel HSI reconstruction model based on INR for mapping spatial coordinates to spectral radiance intensities, and the model builds a bridge between discrete pixels and continuous representation in spectral domain.**
- **The continuous function is approximated by a MLP, whose parameters are predicted by a hypernetwork. Besides, periodic spatial encoding projects the pixel coordinates into a higher dimensional space for recovering more high frequency details.**
- **Experiments on CAVE, NUS, and NTIRE2018 datasets verify the superiority of the model.**

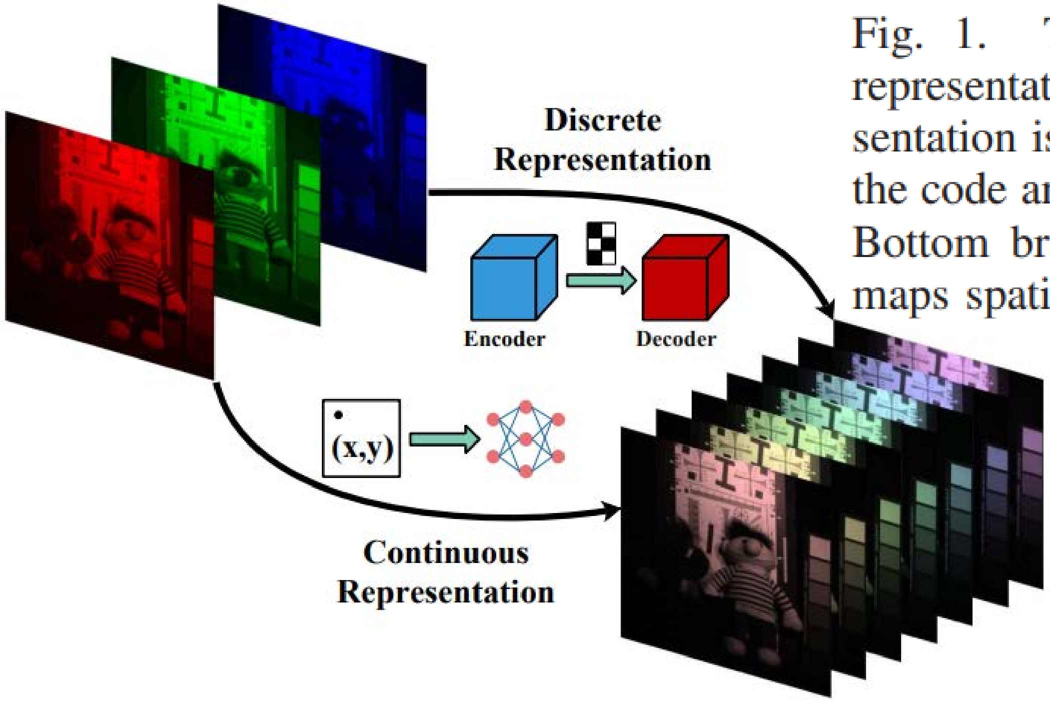
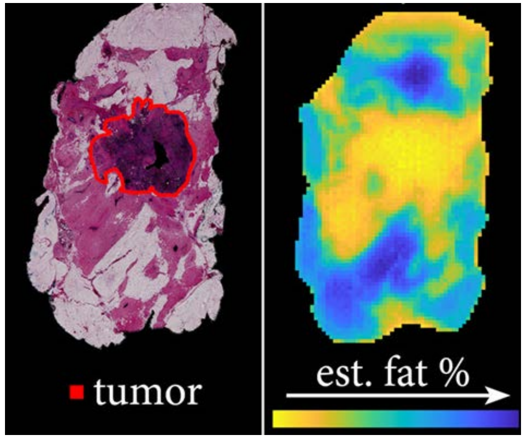
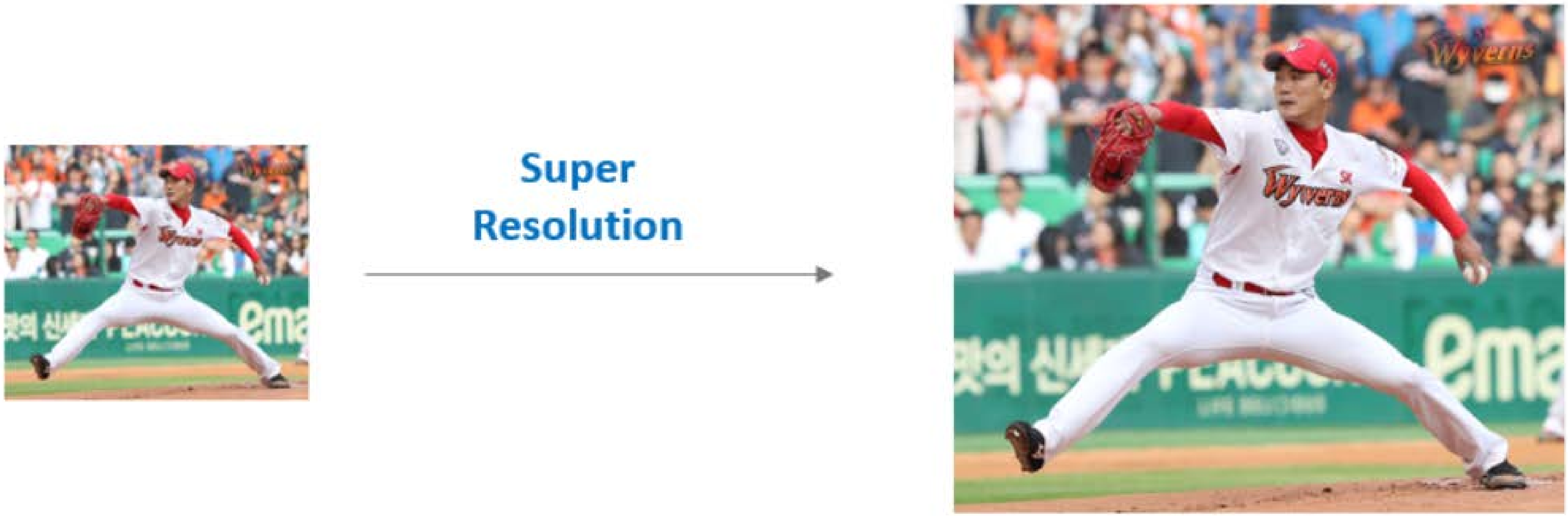


Fig. 1. The hyperspectral image super-resolution process with discrete representation and continuous representation. Top branch: the discrete representation is an autoencoder structure in which the encoder maps the input into the code and the decoder maps the code to a reconstruction of high resolution. Bottom branch: the continuous representation is a continuous function that maps spatial coordinates (x, y) to pixel values.

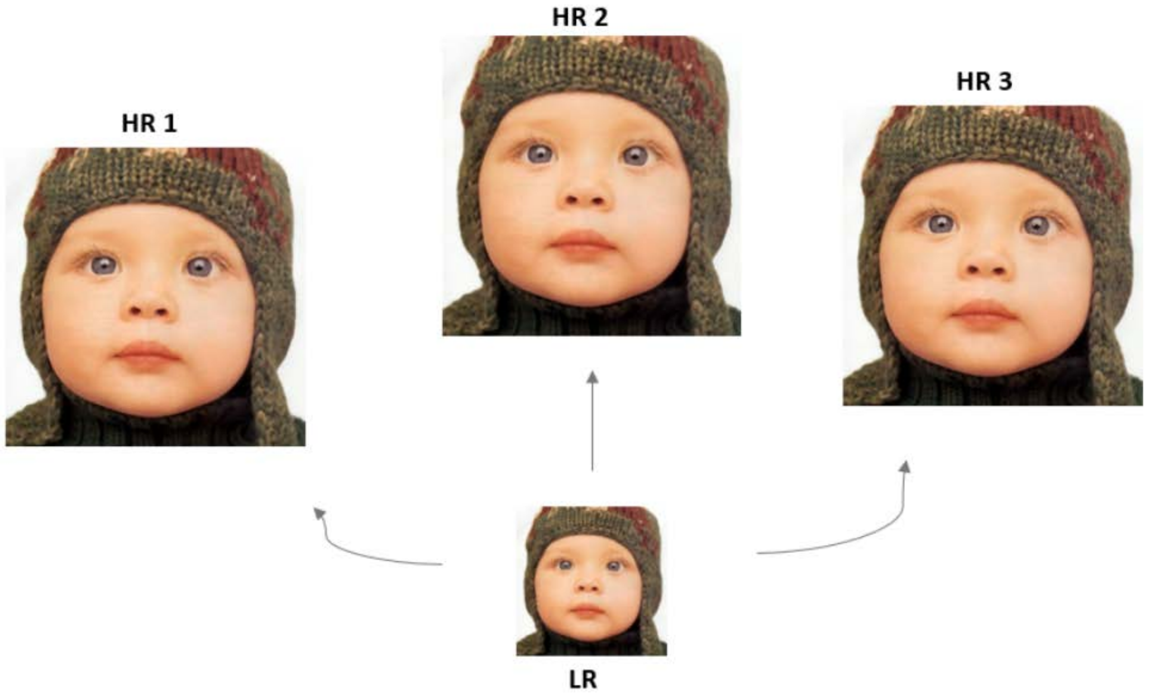


Introduction-Single image super-resolution



[Single Image Super Resolution 예시]

Introduction-Single image super-resolution



[Ill-Posed Problem 예시]

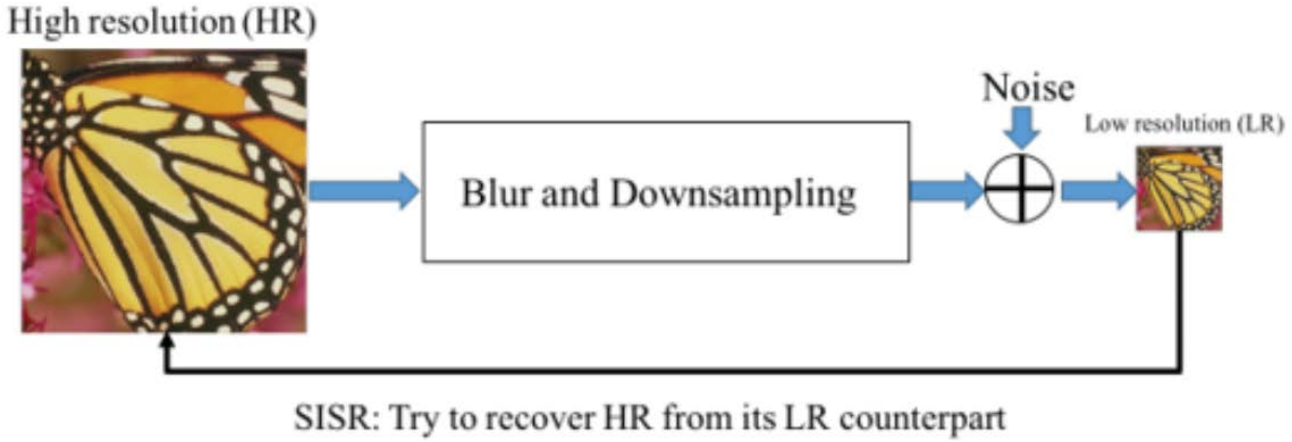


Figure 1: Sketch of the overall framework of SISR.

[Single Image Super Resolution 문제 정의]

Introduction-Why Implicit Neural representation?

Goal: A generative model for 3D-aware image synthesis which allows us to ...

- Generate photorealistic images

- Control individual objects wrt. their pose, size, and position in 3D

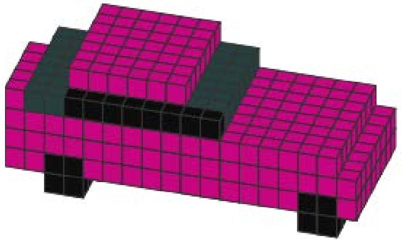
- Control camera viewpoint in 3D

- Train from collections of unposed images

NeRF

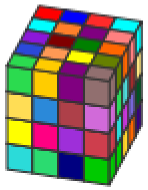
Introduction-NeRF : Representing Scenes as Neural Radiance Fields for View Synthesis

Voxel Representation



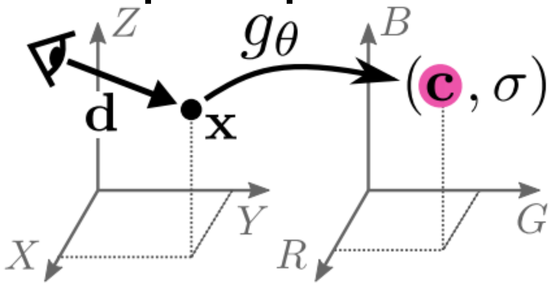
3D Object

Voxel Representation

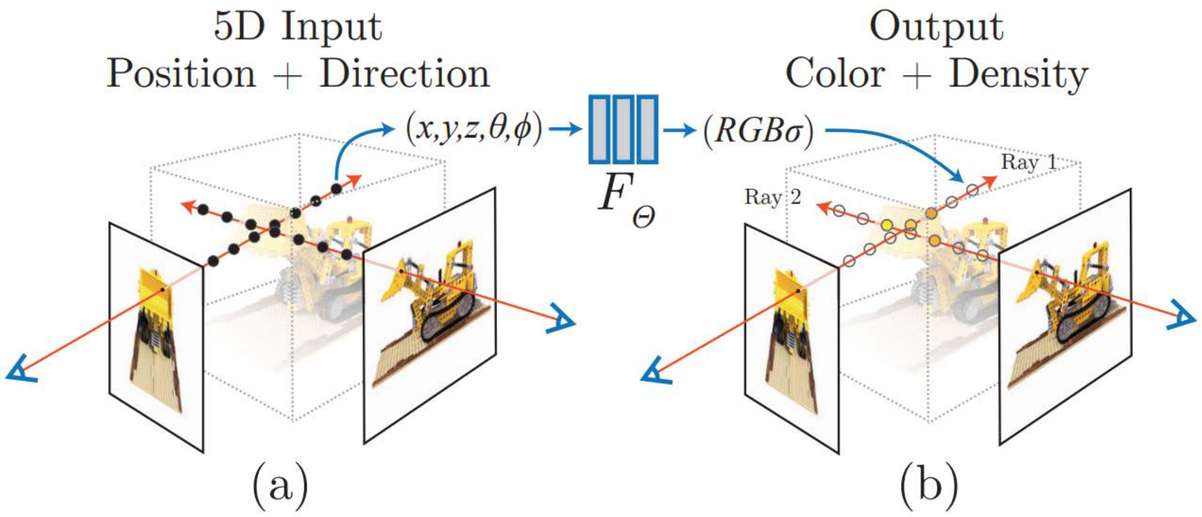


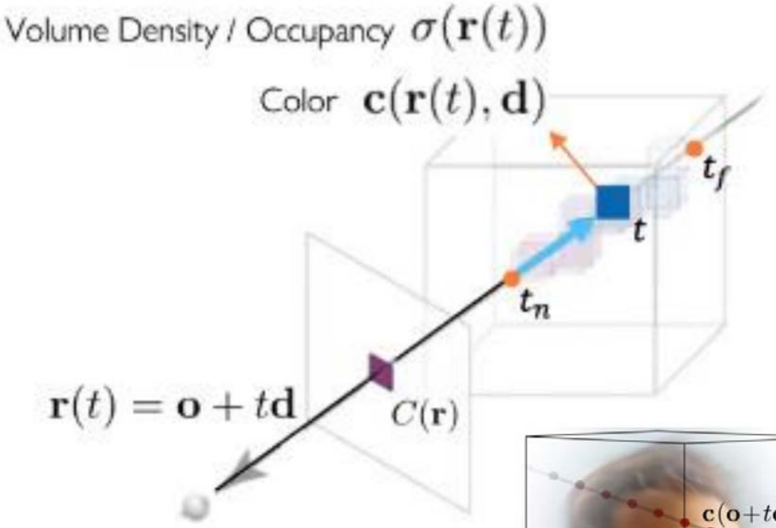
3D Feature

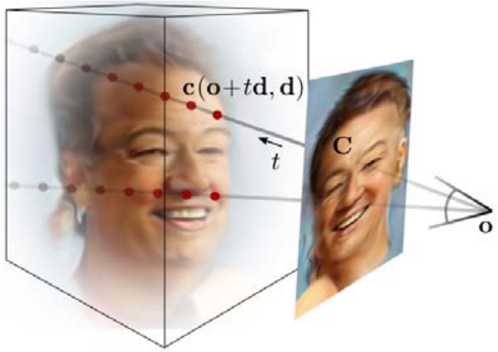
Implicit Representation



Neural Radiance Field





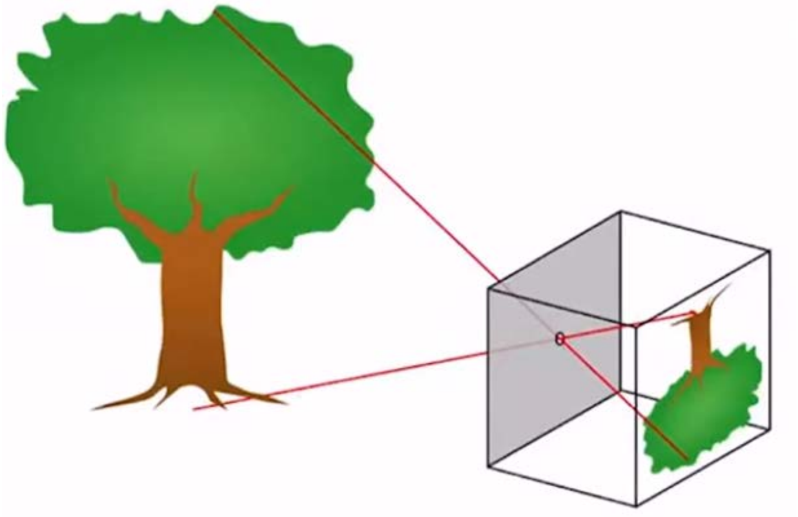


Pixel 값 = 한 Ray 위에 존재하는 point들의 RGB 값들의 Weighted Sum!

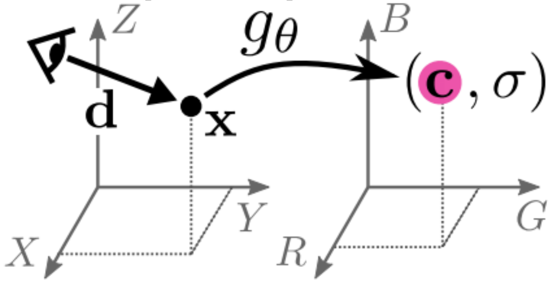
$$C(\mathbf{r}) = \int_{t_n}^{t_f} T(t)\sigma(\mathbf{r}(t))\mathbf{c}(\mathbf{r}(t), \mathbf{d})dt\,, \text{ where } T(t) = \exp\left(-\int_{t_n}^{t} \sigma(\mathbf{r}(s))ds\right)$$

Introduction-NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis

Pinhole Camera Model

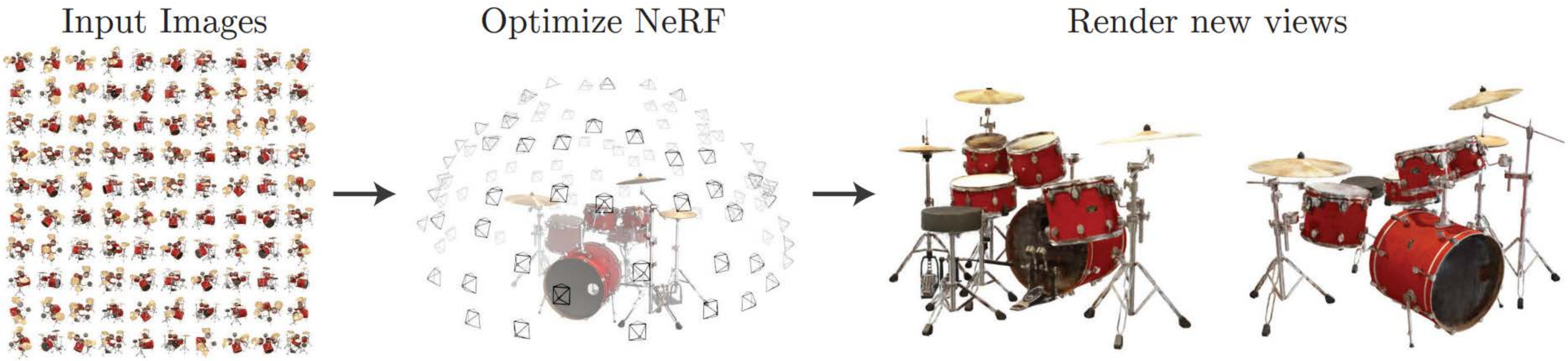


Implicit Representation

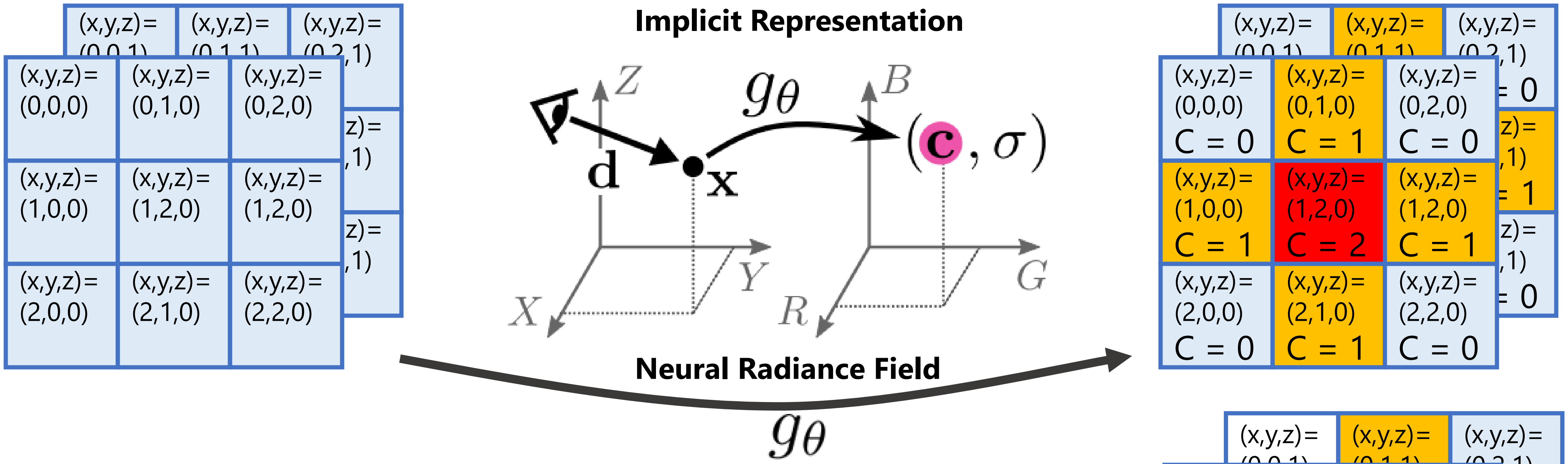


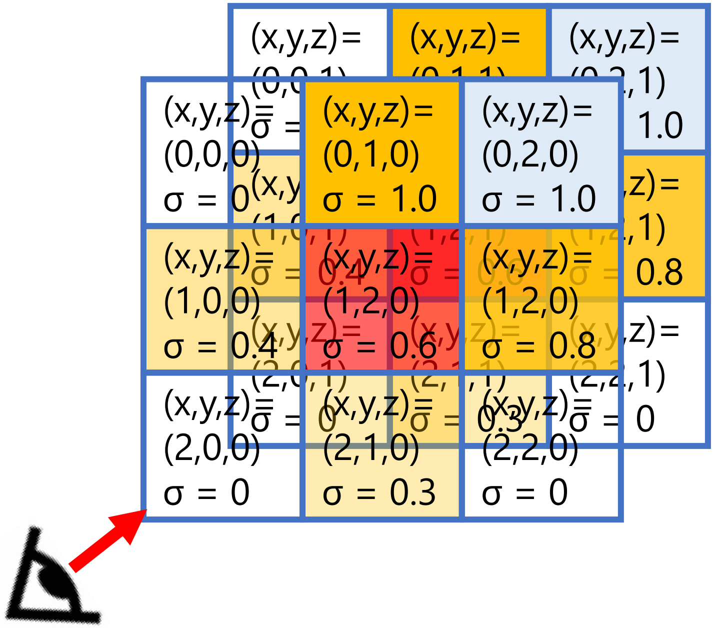
Neural Radiance Field

NeRF



Introduction-Implicit Neural Representation





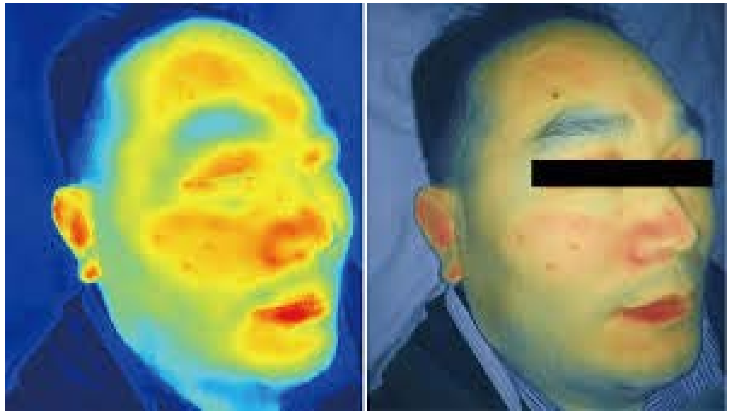
Implicit Representation has the advantages.

- **Continuous representation, multi-view consistent**
- **Lighter than voxel representation (memory consumption), high Image Fidelity**
- **Neural Radiance Field encodes a continuous volume within the deep neural network (MLP), whose input is a single 5D coordinate (x ,y ,z ,θ ,ϕ) and whose output is the volume density and view-dependent RGB color (*r ,g ,b ,σ*(*density*))**

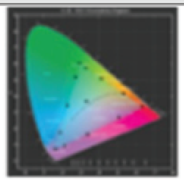
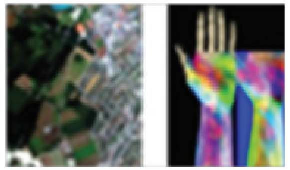
Introduction-Hyperspectral Imaging

분광 이미징

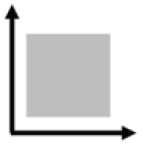
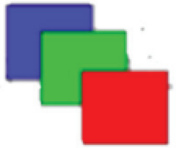
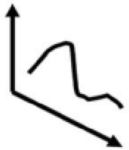
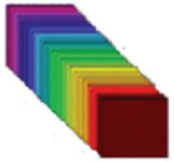
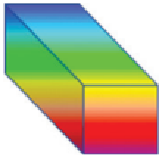
Hyperspectral Imaging



〈표 1〉 초분광 이미징 사용대역

UV	VIS	VIS/NIR	SWIR	MWIR
				
Minerals Forensics Adbasive	Color measurement Printuing works Textile	Medical Food Machine vision Biotechnology Remote sensing Agriculture	Polymers Food Chemicals Military Wood	Chemicals Arts Gas Military

〈표 2〉 밴드수에 따른 분광 기술

	 흑백	 RGB	 분광기	 다분광	 초분광	 극초분광
공간정보	YES	YES	NO	YES	YES	YES
분광 밴드수	1	2	수십~수백	3~10	10~100	〉1000
분광정보	NO	NO	YES	Limited	YES	YES
적용분야	밝기	컬러	고체/액체/가스	고체/액체 탐지	고체/액체 분석	고체/액체/가스 분석
상용화정도	YES	YES	YES	YES	YES	Emerging

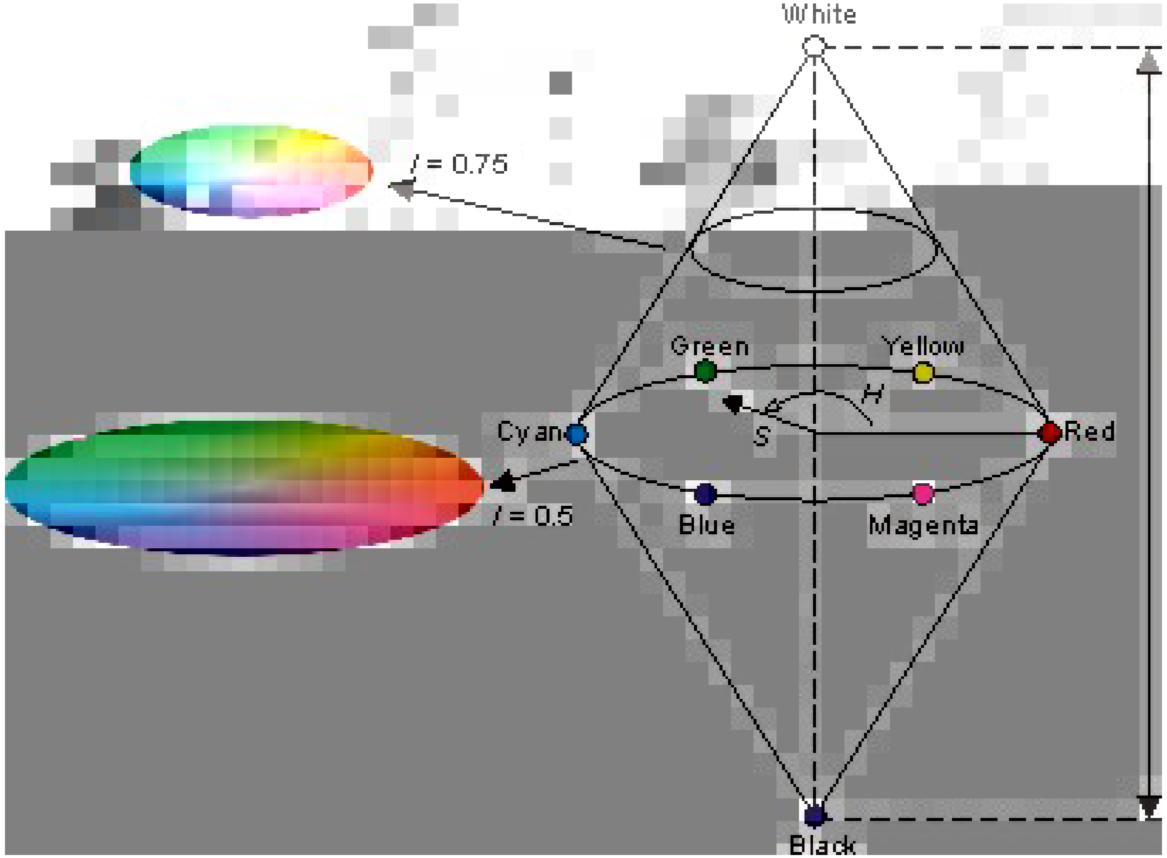
https://ettrends.etri.re.kr/ettrends/175/0905175008/34-1_86-97.pdf

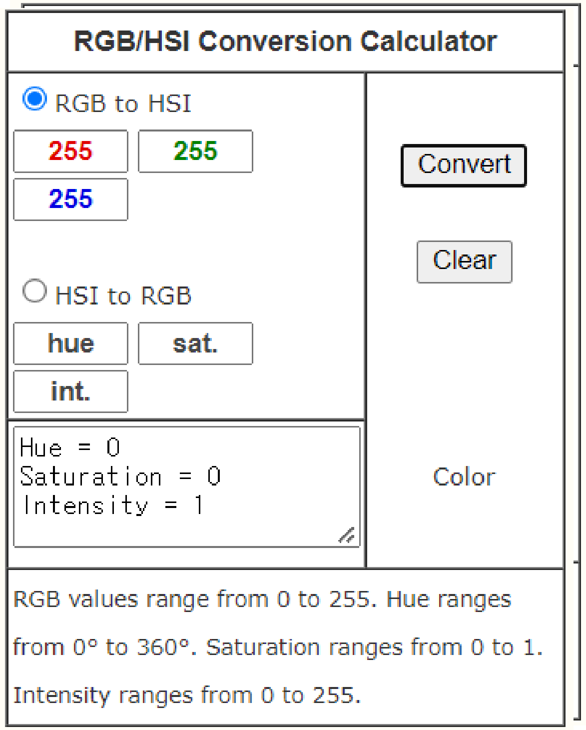
Introduction-HSI color

H (Hue) : 색상

S (Saturation): 채도

I (Intensity): 색상 강도





Methods

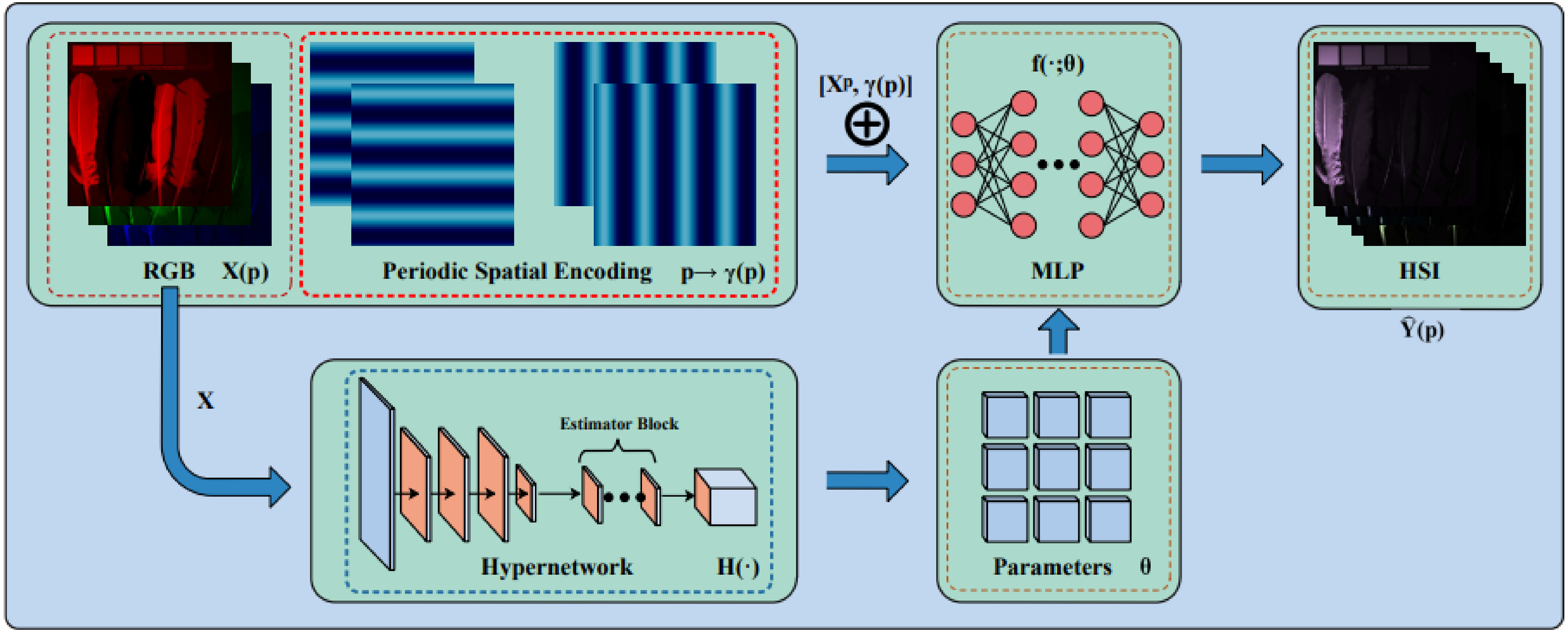
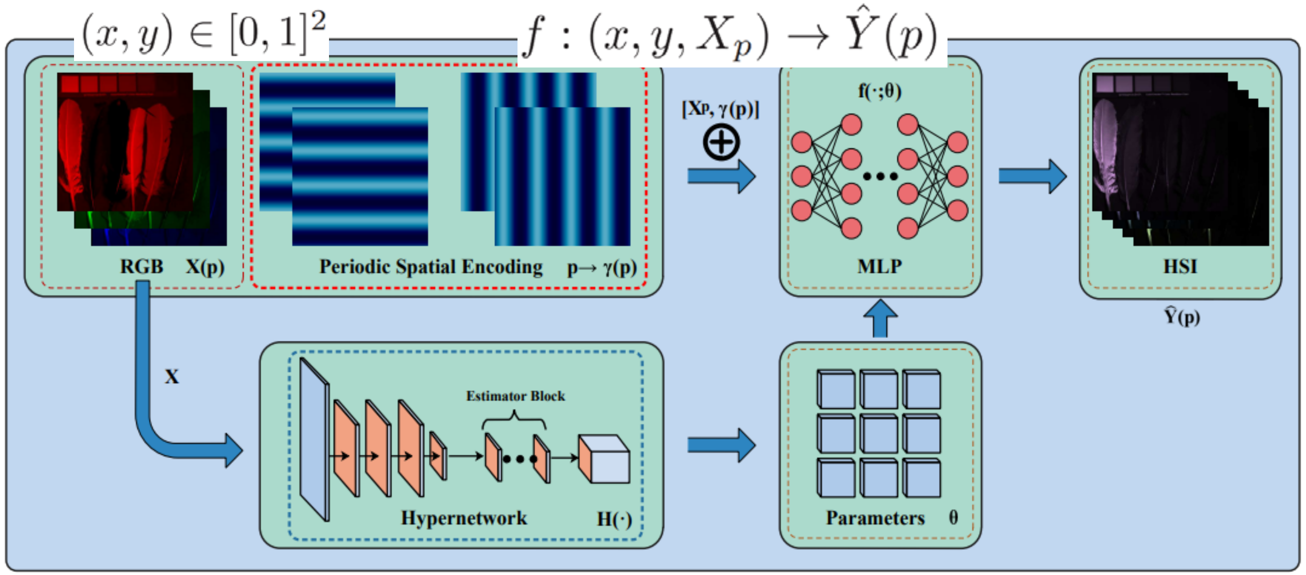


Fig. 2. Architecture overview. Our model first processes the input at hypernetwork to produce a tensor of weights and biases θ for MLP. Then the MLP computes the final output from the input $[X_p, \gamma(p)]$.

Methods – Model Formulation

- $X \in R^{W \times H \times 3}$: **RGB image**
- W, H : **the width, height of an image X**
- $c \in \{R, G, B\}$: **RGB Color 3 channel**
- $Y \in R^{W \times H \times L}$: **ground truth HSI**
- $L\ (L >> 3)$: **the number of spectral bands**
- Λ : **all spectral bands, set**



$$L1 = \|Y - M(X)\|_1 , \quad (1)$$

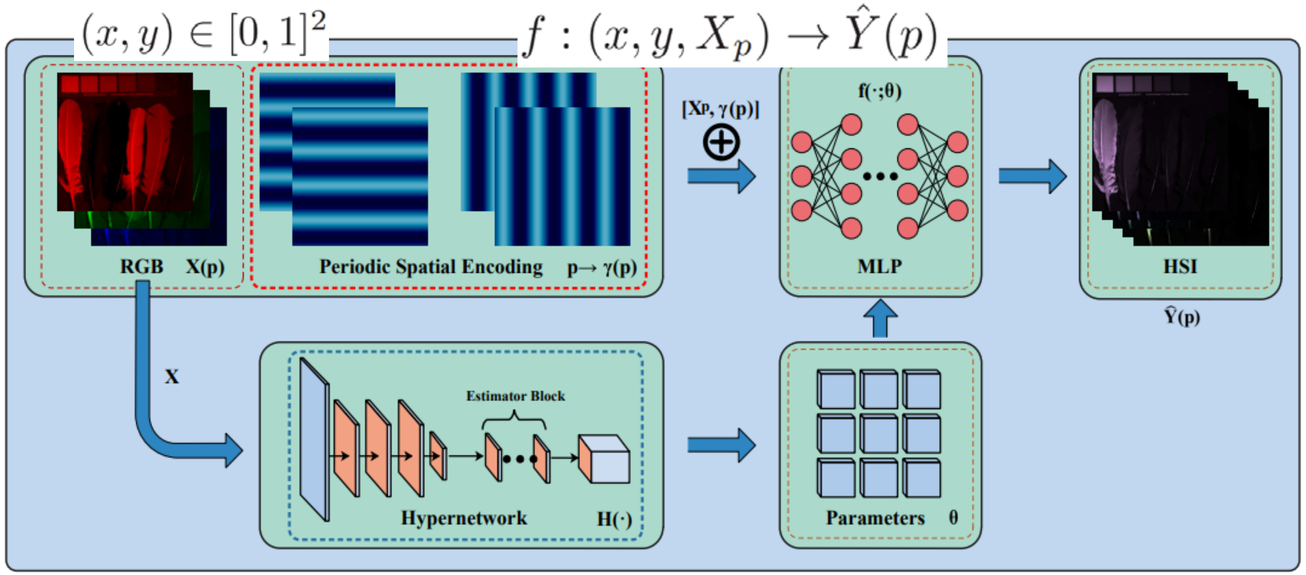
$$Y(p) = \int_{\lambda_1, \ldots, \lambda_L} R(p, \lambda) d\lambda, \quad (2)$$

$$X_c(p) = \int_{\Lambda} R(p, \lambda) \Phi_c(\lambda) d\lambda, \quad (3)$$

$$X_c(p) = \sum_{n=1}^{L} R(p, \lambda_n) \Phi_c(\lambda_n), \quad (4)$$

Methods – Model Formulation

- $X \in R^{W \times H \times 3}$: **RGB image**
- W, H : **the width, height of an image X**
- $c \in \{R, G, B\}$: **RGB Color 3 channel**
- $Y \in R^{W \times H \times L}$: **ground truth HSI**
- $L\ (L >> 3)$: **the number of spectral bands**
- Λ : **all spectral bands, set**



$$L1 = \|Y - M(X)\|_1 , \qquad (1)$$

$$Y(p) = \int_{\lambda_1, \ldots, \lambda_L} R(p, \lambda) d\lambda, \quad (2)$$

$$X_c(p) = \int_{\Lambda} R(p, \lambda) \Phi_c(\lambda) d\lambda, \quad (3)$$

$$X_c(p) = \sum_{n=1}^{L} R(p, \lambda_n) \Phi_c(\lambda_n), \quad (4)$$

$$X_{c=R} = \int R(p, \lambda)\, \Phi_{c=R}(\lambda) d\lambda$$

$$X_{c=G} = \int R(p, \lambda)\, \Phi_{c=G}(\lambda) d\lambda$$

$$X_{c=B} = \int R(p, \lambda)\, \Phi_{c=B}(\lambda) d\lambda$$

Methods – Model Formulation

- $X \in R^{W \times H \times 3}$ **: RGB image**
- W, H **: the width, height of an image X**
- $c \in \{R, G, B\}$ **: RGB Color 3 channel**
- $Y \in R^{W \times H \times L}$ **: ground truth HSI corresponding to X**
- $L\ (L >> 3)$ **: the number of spectral bands in hyperspectral images**
- Λ **: all spectral bands, set**

$$X_{c=R} = \int R(p,\lambda)\,\Phi_{c=R}(\lambda)d\lambda$$

$$X_{c=G} = \int R(p,\lambda)\,\Phi_{c=G}(\lambda)d\lambda$$

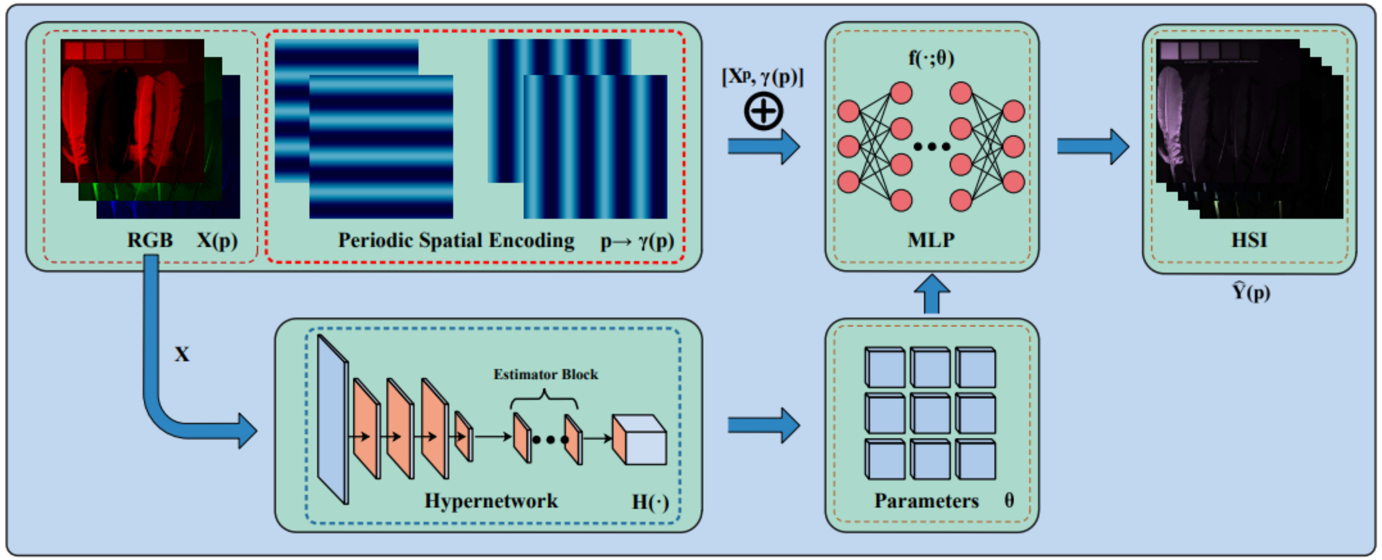
$$X_{c=B} = \int R(p,\lambda)\,\Phi_{c=B}(\lambda)d\lambda$$

$$L1 = \|Y - M(X)\|_1\,, \qquad (1)$$

$$Y(p) = \int_{\lambda_1,\ldots,\lambda_L} R(p,\lambda)d\lambda, \ (2)$$

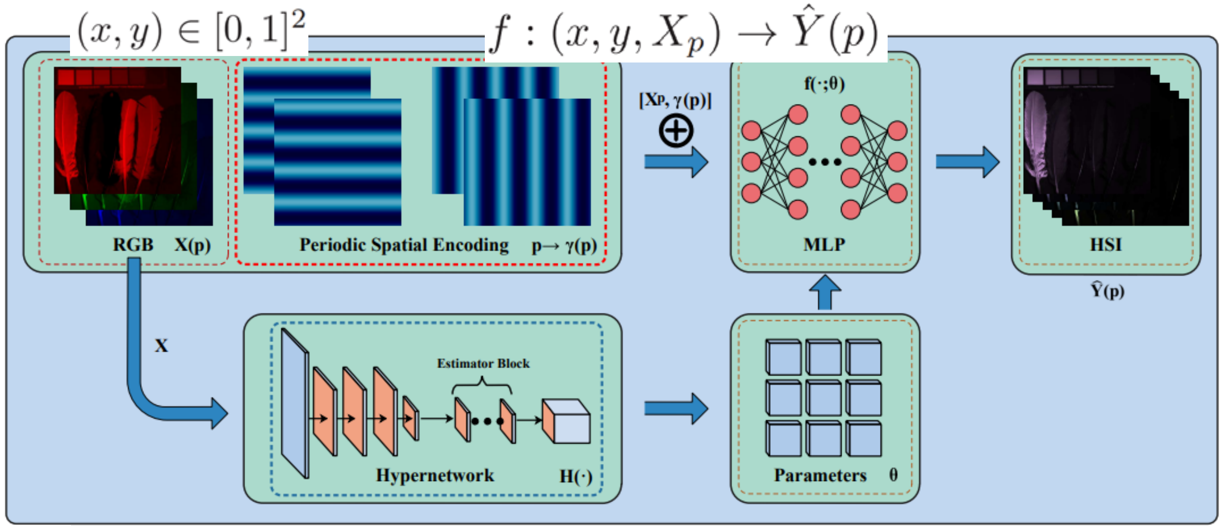
$$X_c(p) = \int_{\Lambda} R(p,\lambda)\Phi_c(\lambda)d\lambda, \ (3)$$

$$X_c(p) = \sum_{n=1}^{L} R(p,\lambda_n)\Phi_c(\lambda_n), \ (4)$$



Methods – Periodic Spatial Encoding

- $X \in R^{W \times H \times 3}$: **RGB image**
- W, H : **the width, height of an image X**
- $c \in \{R, G, B\}$: **RGB Color 3 channel**
- $Y \in R^{W \times H \times L}$: **ground truth HSI**
- $L\ (L >> 3)$: **the number of spectral bands**
- Λ : **all spectral bands, set**
- $\gamma_k(p)$: **periodic spatial encoding from** R^2 to R^4



$$f(X_p, p; \theta) = \hat{Y}(p), \quad (5)$$

$$\theta = H(X), \quad (6)$$

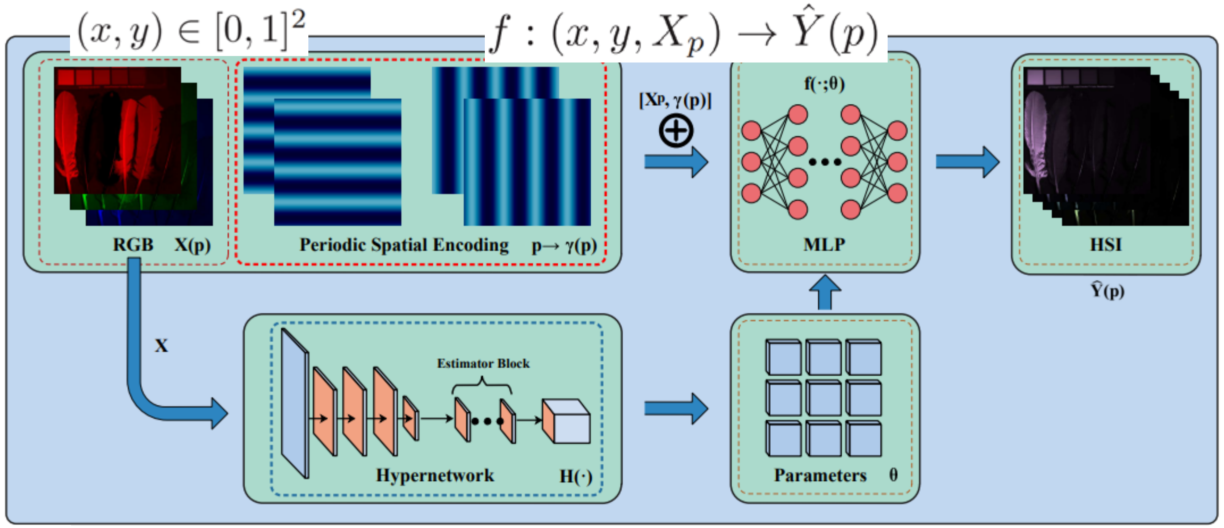
$$\gamma_k(p) = [\cos(2^k \pi x), \sin(2^k \pi x), \cos(2^k \pi y), \sin(2^k \pi y)]. \quad (7)$$ **:4 channel**

$$\gamma(p) = [\gamma_0(p), \ldots, \gamma_{N-1}(p)]. \quad (8) \qquad R^{4N}\ (4N > 2)$$ **:4N channel**

$$f(X_p, \gamma(p); \theta) = \hat{Y}(p). \quad (9) \qquad R^2 \rightarrow R^{4N+3}$$ X_p **:3 channel**

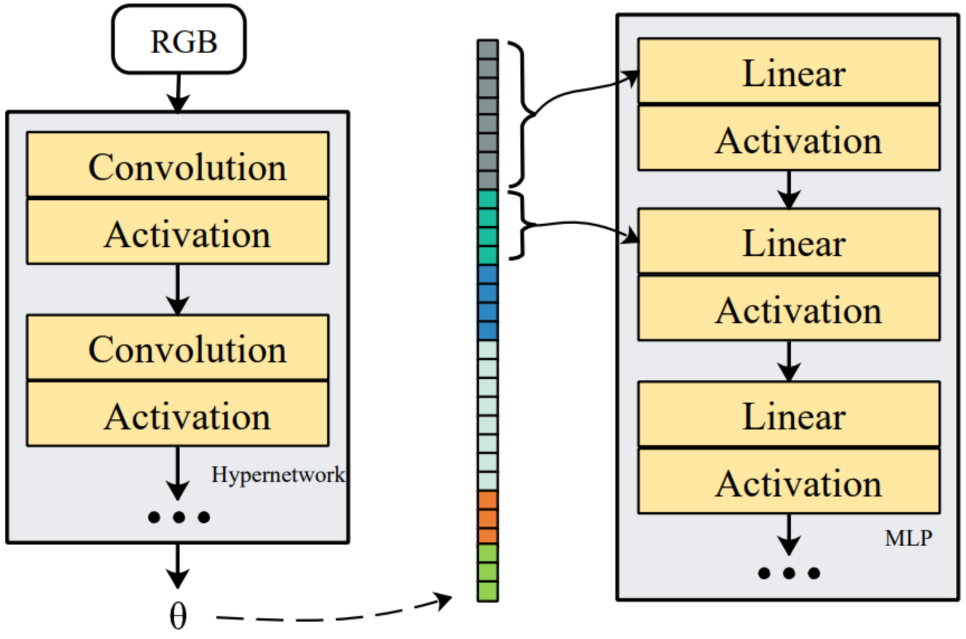
Methods – Parametric Model of Implicit Neural Representation

- $X \in R^{W \times H \times 3}$: **RGB image**
- W, H : **the width, height of an image X**
- $c \in \{R, G, B\}$: **RGB Color 3 channel**
- $Y \in R^{W \times H \times L}$: **ground truth HSI**
- $L \ (L >> 3)$: **the number of spectral bands**
- Λ : **all spectral bands, set**
- $\gamma_k(p)$: **periodic spatial encoding from** R^2 to R^4
- $\theta = H(X)$



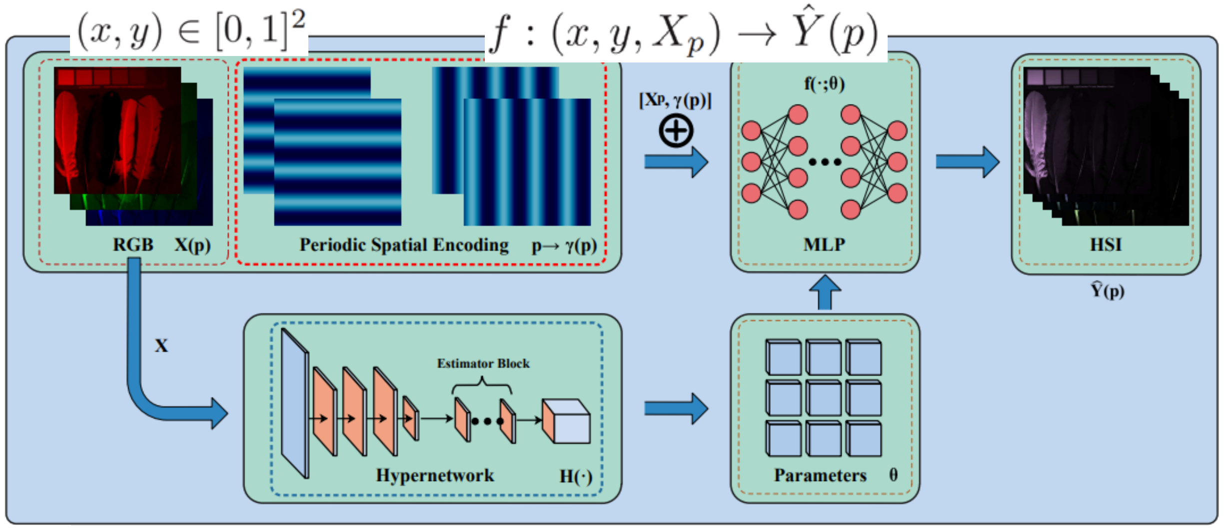
The model parameterization method enables realistic image output with consistency edge and detailed structures

$$\theta = [\theta_{cell_i}]_{i=1,\ldots,S \times S}.$$



Methods

- $X \in R^{W \times H \times 3}$: **RGB image**
- W, H : **the width, height of an image X**
- $c \in \{R, G, B\}$: **RGB Color 3 channel**
- $Y \in R^{W \times H \times L}$: **ground truth HSI**
- $L\ (L >> 3)$: **the number of spectral bands**
- Λ : **all spectral bands, set**
- $\gamma_k(p)$: **periodic spatial encoding from** R^2 to R^4
- $\theta = H(X)$



$$\theta = \arg\min_{\theta} \sum^{p \in R^{W*H}} |f(X_p, \gamma(p); \theta) - Y(p)|, \qquad (11)$$

Experiments

Datasets

- **Cave**
- **NUS**
- **NTIRE2018**

Evaluation Metric

$$PSNR(Y,\hat{Y}) = 10log_{10}(\frac{max(Y)^2}{\frac{1}{W \times H}\left\|Y-\hat{Y}\right\|_2^2}). \quad (12)$$

$$SSIM(Y,\hat{Y}) = \frac{(2\mu_Y\mu_{\hat{Y}}+c_1)(2\sigma_{Y\hat{Y}}+c_2)}{(\mu_Y^2+\mu_{\hat{Y}}^2+c_1)(\sigma_Y^2+\sigma_{\hat{Y}}^2+c_2)}, \quad (13)$$

$$SAM(Y,\hat{Y}) = arccos\frac{<Y^T\hat{Y}>}{\|Y\|_2 \cdot \left\|\hat{Y}\right\|_2}. \quad (14)$$

Experiments

Datasets

- **Cave**
- **NUS**
- **NTIRE2018**

Evaluation Metric

$$PSNR(Y,\hat{Y}) = 10log_{10}(\frac{max(Y)^2}{\frac{1}{W\times H}\left\|Y-\hat{Y}\right\|_2^2}). \quad (12)$$

$$l(A,B) = \frac{2\mu_A\mu_B + C_1}{\mu_A^2 + \mu_B^2 + C_1}$$

휘도를 비교 하기 위한 수식

$$SSIM(Y,\hat{Y}) = \frac{(2\mu_Y\mu_{\hat{Y}} + c_1)(2\sigma_{Y\hat{Y}} + c_2)}{(\mu_Y^2 + \mu_{\hat{Y}}^2 + c_1)(\sigma_Y^2 + \sigma_{\hat{Y}}^2 + c_2)}, \quad (13)$$

$$c(A,B) = \frac{2\sigma_A\sigma_B + C_2}{\sigma_A^2 + \sigma_B^2 + C_2}$$

대비를 비교 하기 위한 수식

$$SAM(Y,\hat{Y}) = arccos\frac{< Y^T\hat{Y} >}{\|Y\|_2 \cdot \left\|\hat{Y}\right\|_2}. \quad (14)$$

$$S(A,B) = \frac{\sigma_{AB} + C_3}{\sigma_A\sigma_B + C_3}$$

구조 비교를 위한 식

$$SSIM(A,B) = l(A,B)c(A,B)s(A,B)$$

Experiments – CAVE Dataset

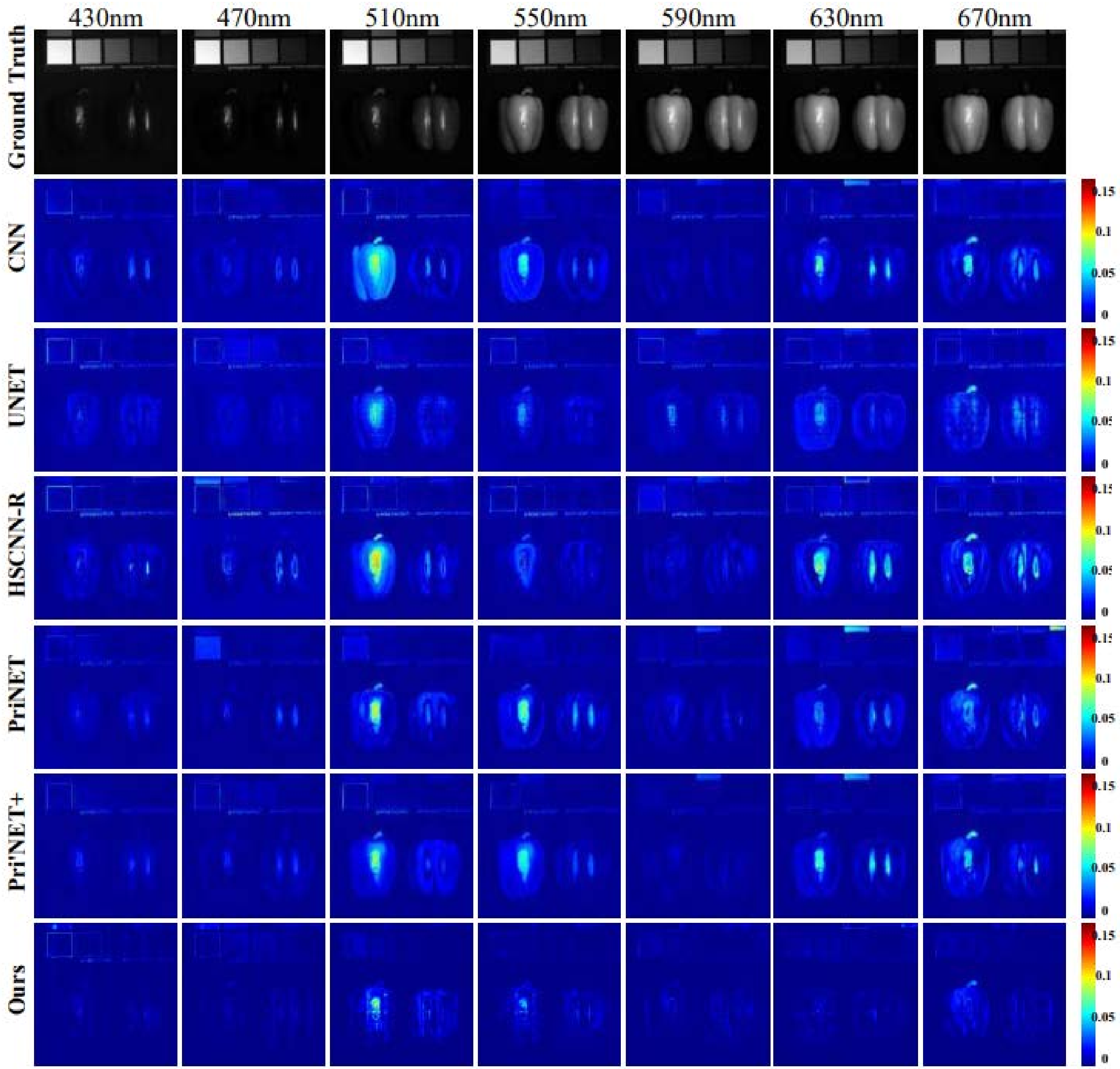
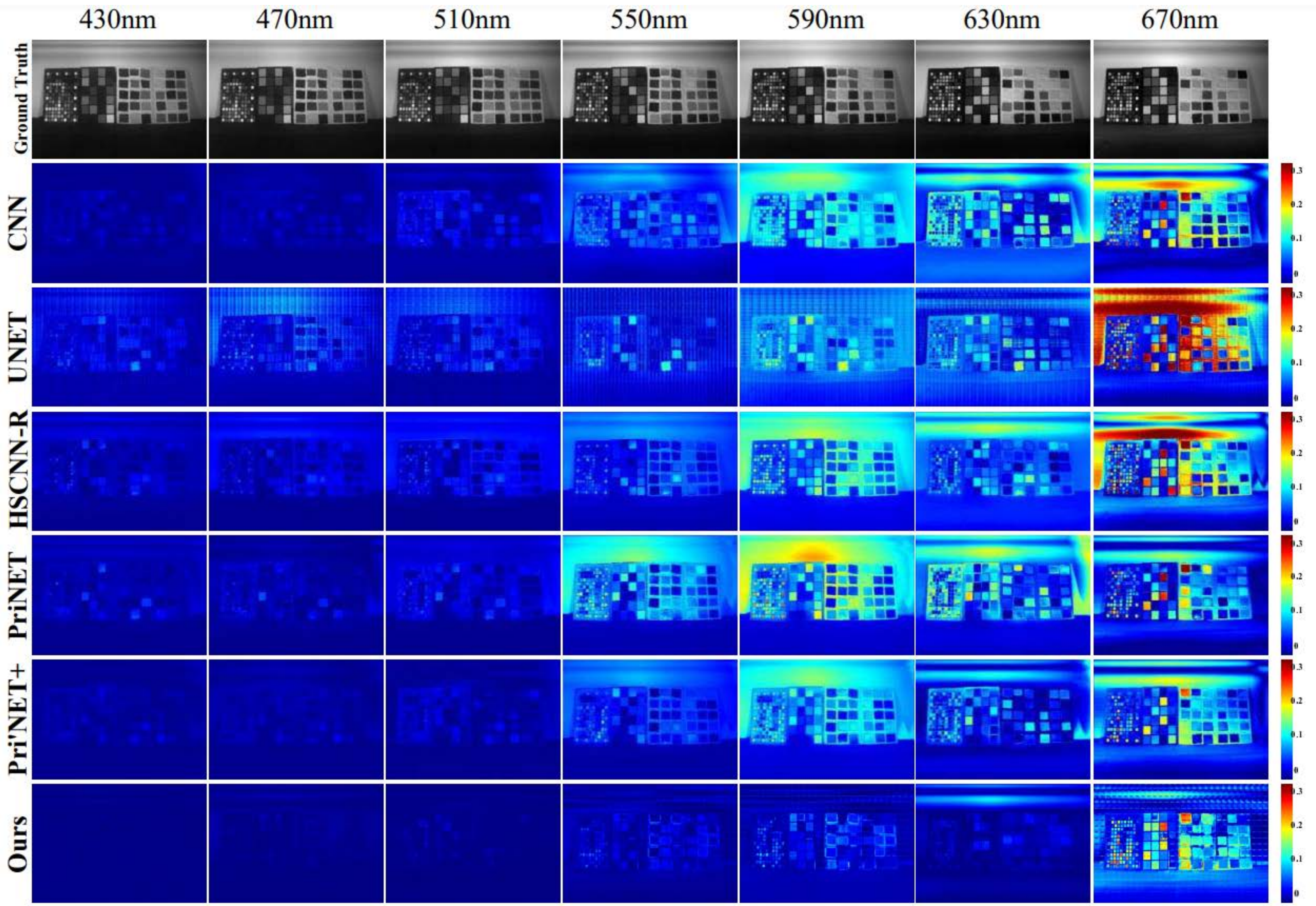


TABLE I
QUANTITATIVE RESULTS OF DIFFERENT METHODS ON CAVE, NUS, AND NTIRE2018 DATASETS.

Model	CAVE			NUS			NTIRE2018		
	PSNR	SSIM	SAM	PSNR	SSIM	SAM	PSNR	SSIM	SAM
CNN [51]	32.2165	0.9706	10.7111	25.5296	0.9238	9.4873	45.8232	0.9998	1.7232
UNET [52]	31.6973	0.9488	13.1052	25.0038	0.8872	10.0554	38.2989	0.9976	3.0454
HSCNN-R [53]	31.4676	0.9637	12.2081	25.1326	0.9213	9.5254	45.7062	0.9998	1.6455
Multi-scale CNN [54]	31.9298	0.9575	11.8746	25.1922	0.9219	9.5021	45.7752	0.9998	1.6938
PriNET [55]	32.8129	0.9733	10.0400	25.2622	0.9368	9.9859	46.2661	0.9999	1.5560
PriNET+ [35]	32.8300	**0.9833**	8.7750	26.2893	0.9405	8.9923	**46.3500**	0.9999	1.5316
INR(Ours)	**34.6257**	0.9781	**7.3285**	**26.3431**	**0.9568**	**8.8187**	46.0808	**0.9999**	**1.5198**

Experiments - NUS Dataset



Ablation Study

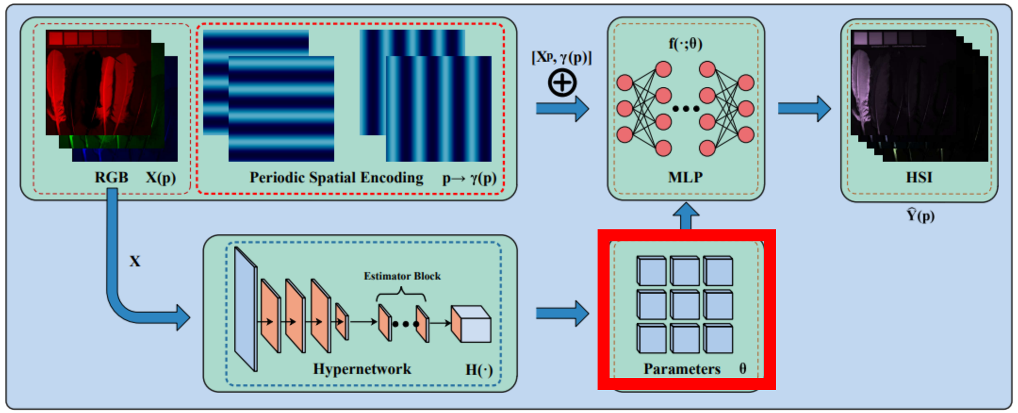
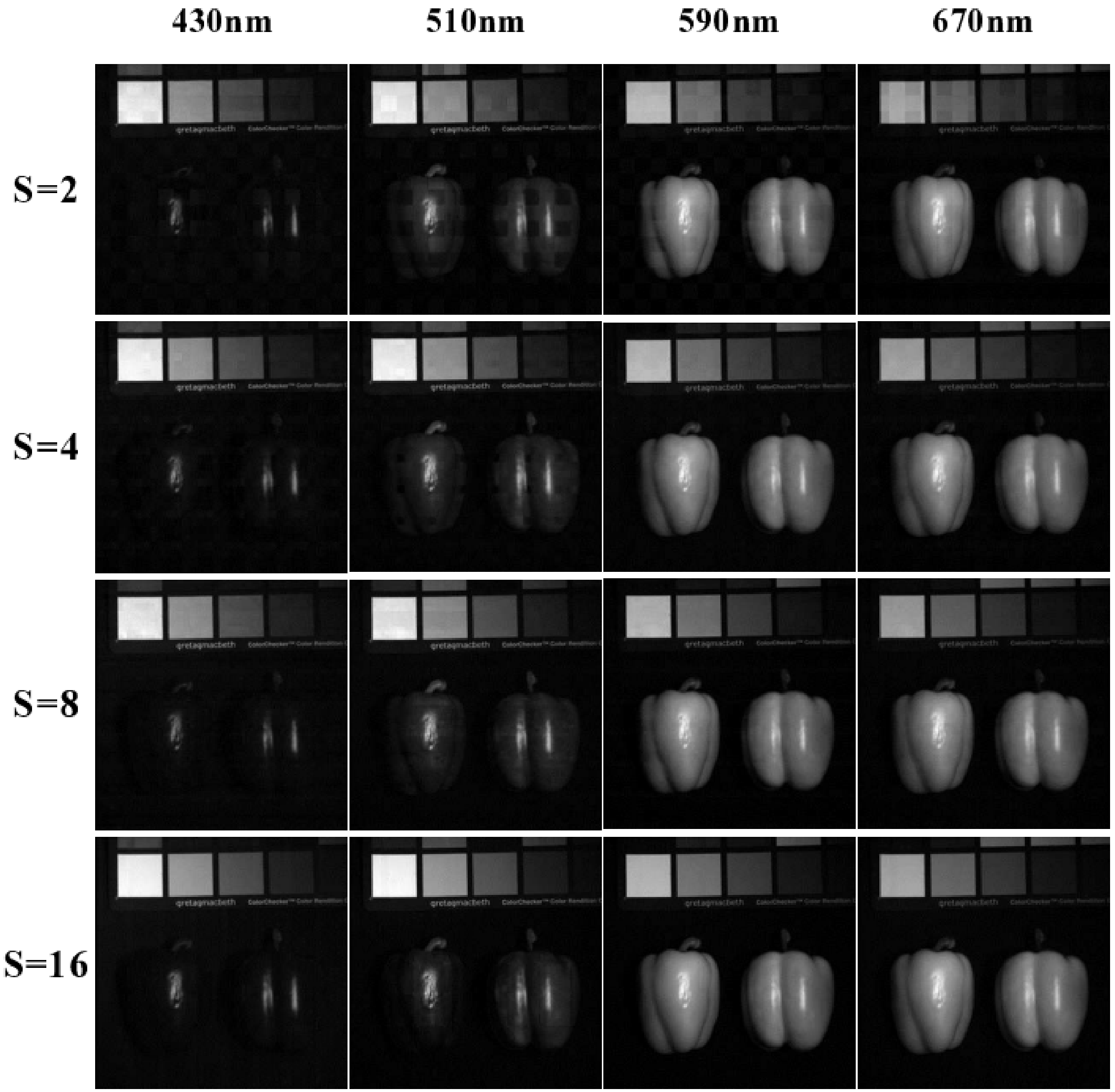


TABLE II
THE RECONSTRUCTION RESULTS ON DIFFERENT SIZES OF PARAMETER GRID.

S	#Ds layers	Params	PSNR	SSIM	SAM
S=2	5	64.9M	**33.8938**	0.9718	7.5979
S=4	4	64.9M	33.1491	0.9729	7.5561
S=8	3	64.8M	33.4025	**0.9730**	7.6563
S=16	2	64.8M	33.8625	0.9702	**7.5359**

TABLE III
RECONSTRUCTION RESULTS ON DIFFERENT FREQUENCIES OF PERIODIC SPATIAL ENCODING.

Per. Spa. Enc.	S	PSNR	SSIM	SAM
w/o	S=16	32.6752	0.9669	8.3312
N=3	S=16	34.0027	0.9747	7.7469
N=5	S=16	**34.6257**	**0.9781**	**7.3285**



Q&A