# Cascaded Local Implicit Transformer for Arbitrary-Scale Super-Resolution (CLIT)

**Hao-Wei Chen**[*,1,2], **Yu-Syuan Xu**[*,1,2], Min-Fong Hong[2], Yi-Min Tsai[2], Hsien-Kai Kuo[2], and Chun-Yi Lee[1]

*Equal contribution [1]Elsa Lab, National Tsing Hua University [2]MediaTek Inc.

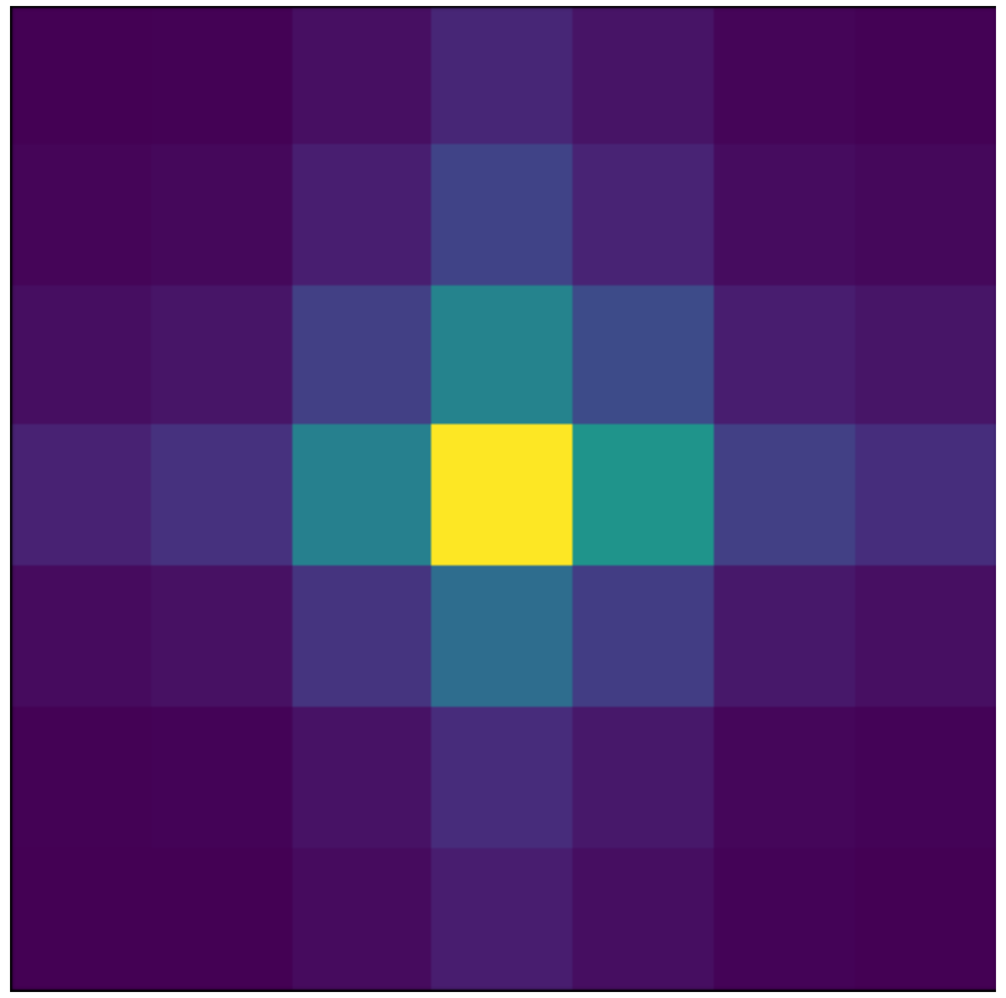
**THU-AM-170**

JUNE 18-22, 2023
CVPR
VANCOUVER, CANADA

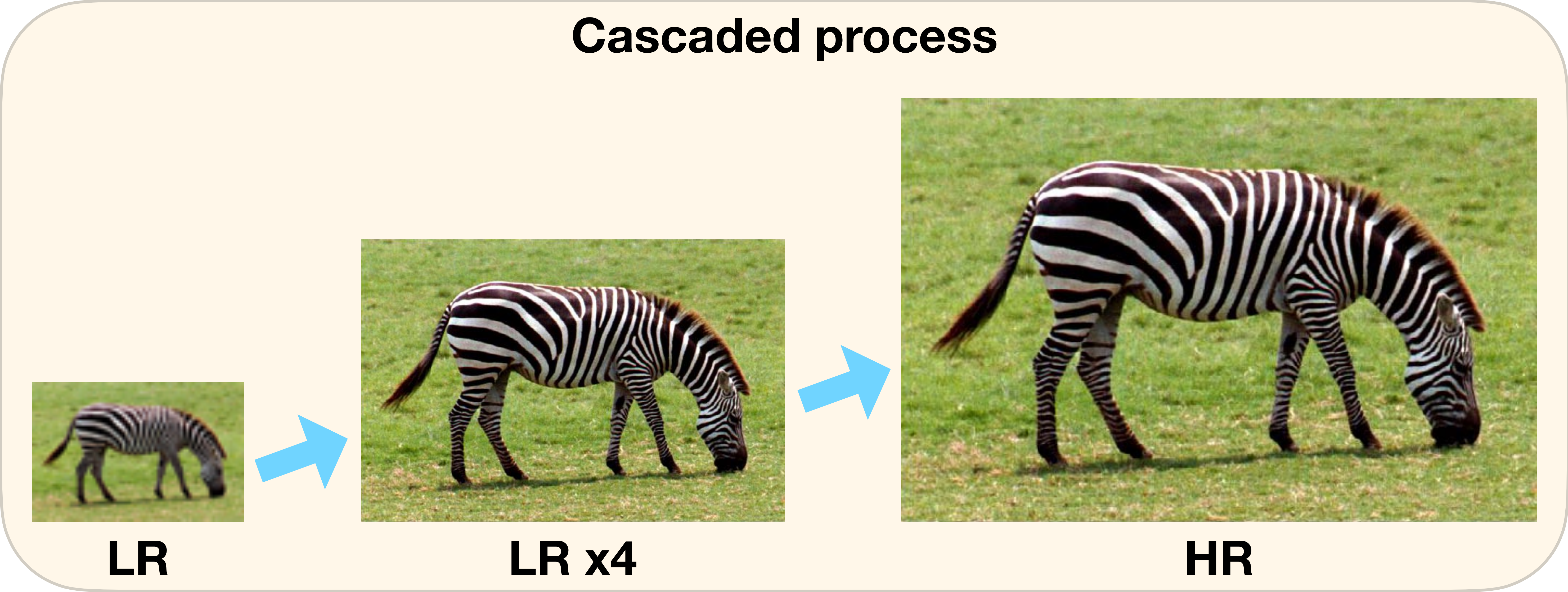
# Summary and Contributions

- ✴ Super-Resolution (SR)
- ✴ High-Resolution (HR)
- ✴ Low-Resolution (LR)

- **Local Implicit Transformer (LIT) for arbitrary-scale SR**
  - Introduce the concept of attention into the arbitrary-scale SR
- **Cascaded LIT (CLIT) for further enhancing performance**
  - A cascaded framework for progressively upscaling LR images
  - CLIT employs a cumulative training strategy



Local attention map



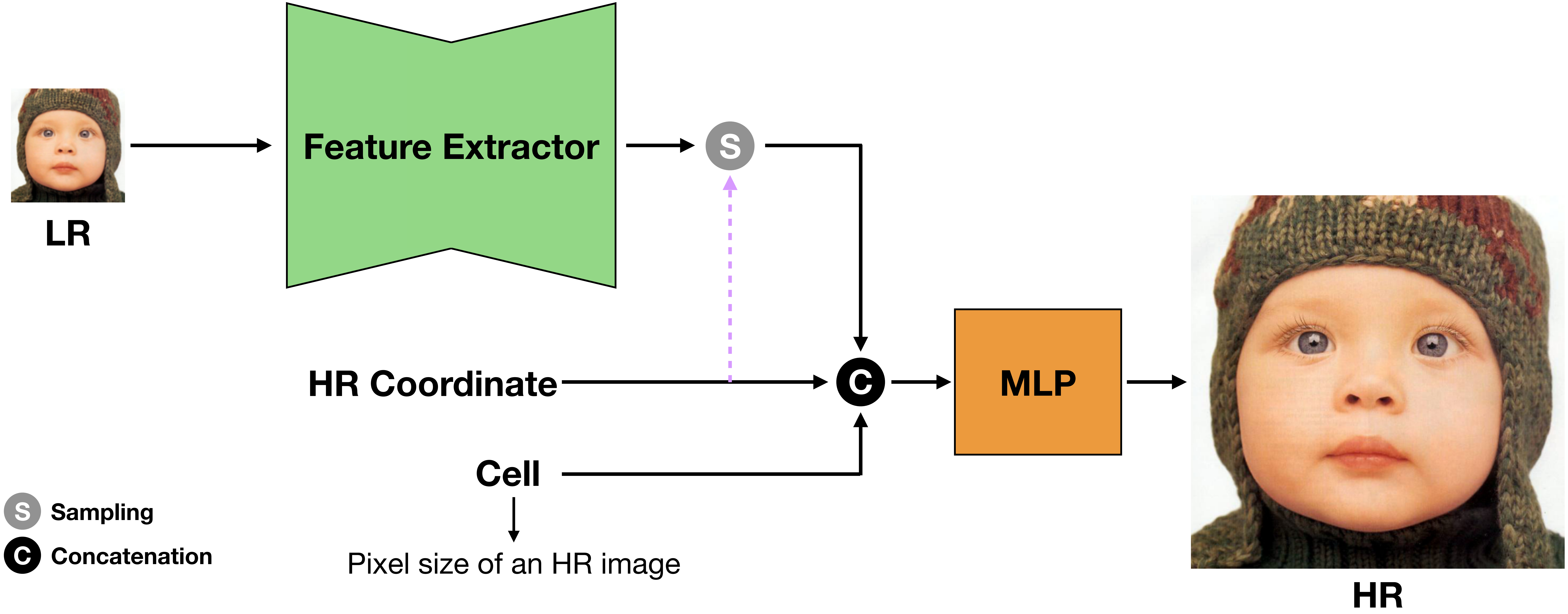
Cascaded process

# Background

# Background

## Arbitrary-scale SR

### LIIF [1] borrows the concept from neural implicit function



LR

Feature Extractor

S

HR Coordinate

Cell

Pixel size of an HR image

C

MLP

HR

S Sampling

C Concatenation

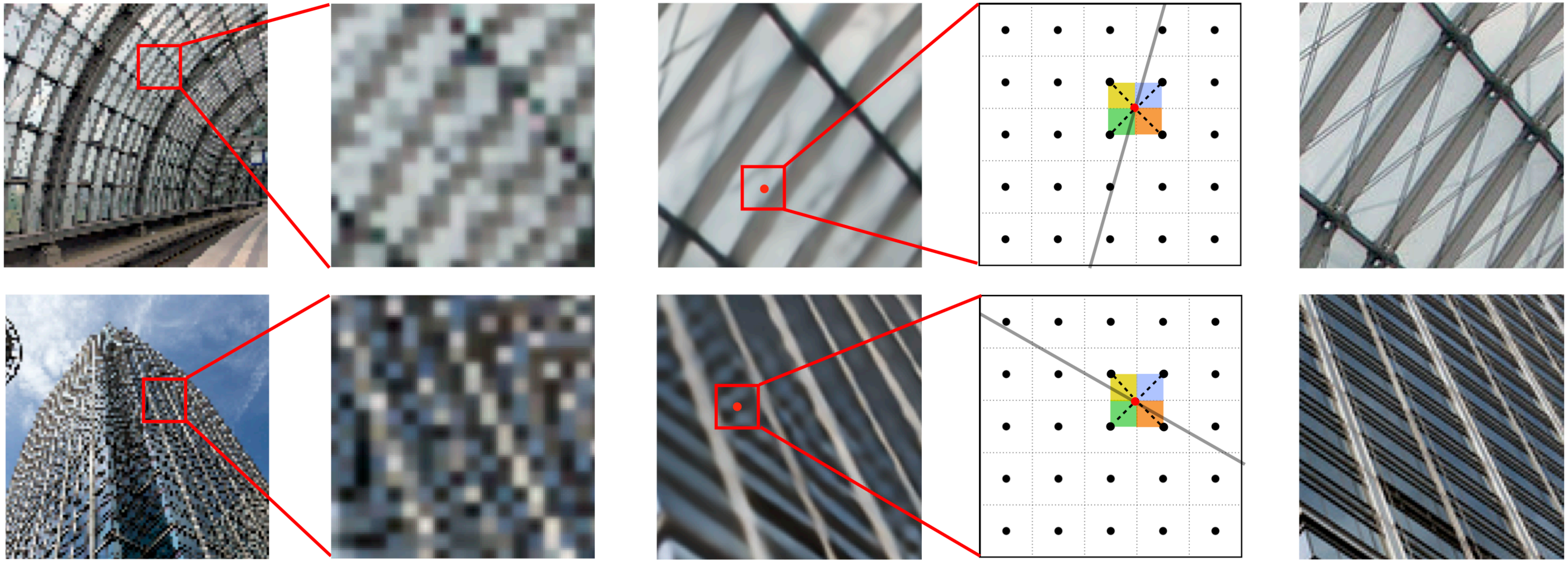
[1] Y. Chen *et al.*, Learning continuous image representation with local implicit image function, CVPR 2021.

# Motivation

# Motivation

## Local ensemble (bilinear interpolation)

**The RGB value of a queried coordinate is calculated by the weighted average of its surrounding four pixels based only on the relative distances**



Input LR LR LIIF [1] Local Ensemble HR

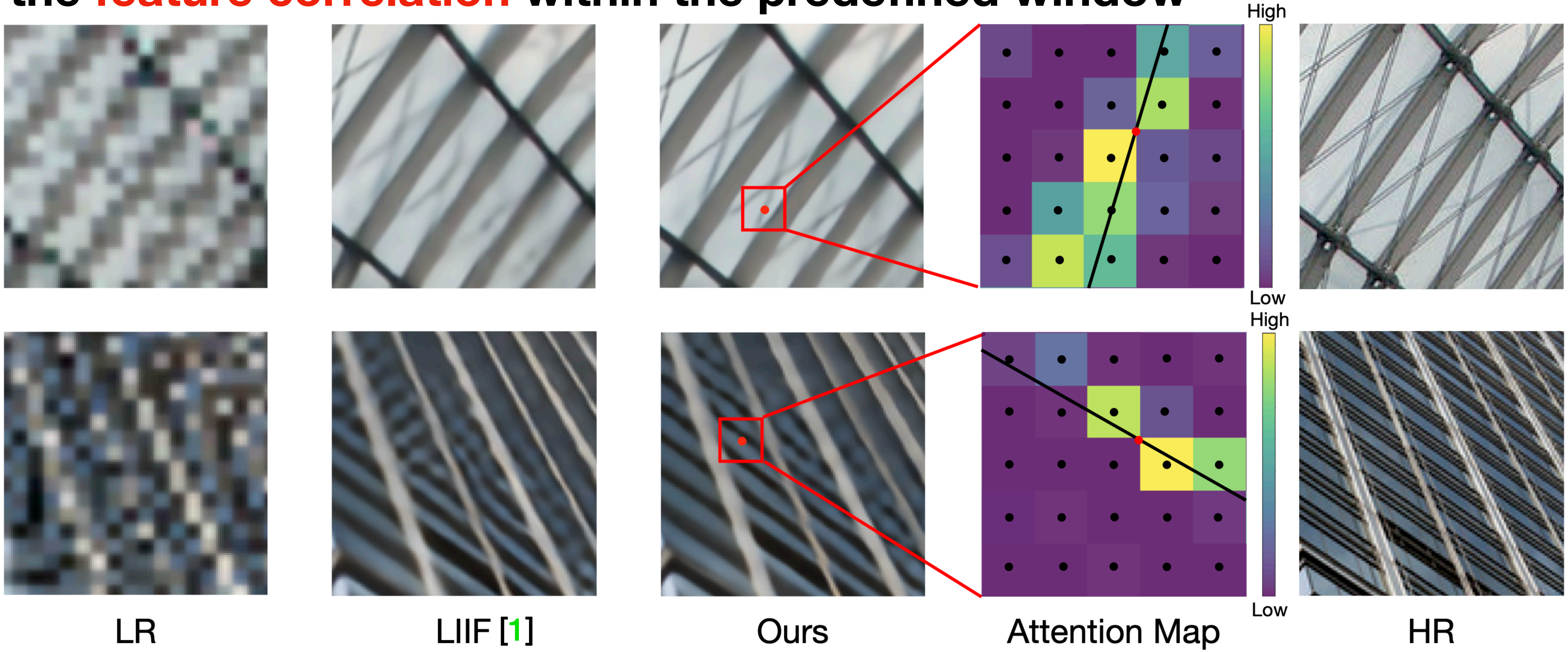
[1] Y. Chen *et al.*, Learning continuous image representation with local implicit image function, CVPR 2021.

[1] Y. Chen *et al.*, Learning continuous image representation with local implicit image function, CVPR 2021.

# Motivation

## Local attention mechanism

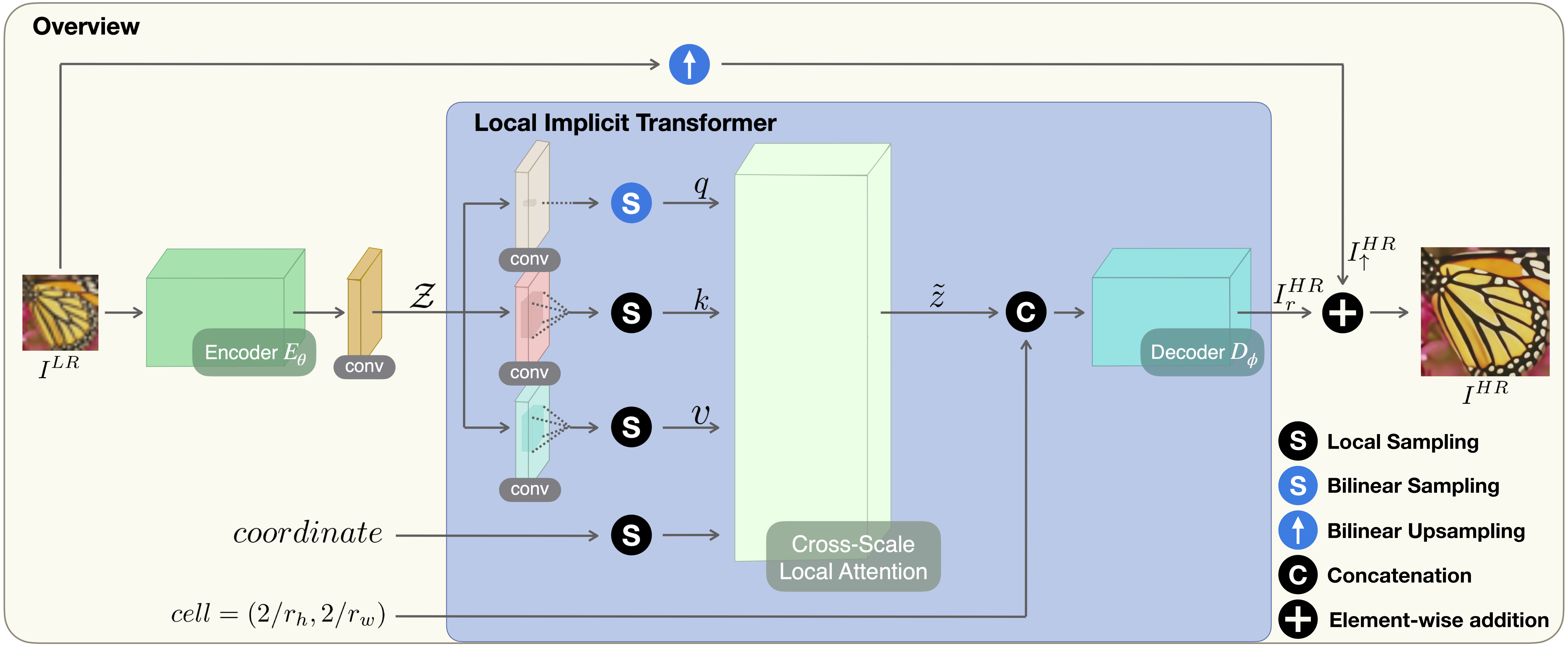
**Expand the referenced area and exploit the attention mechanism to account for the feature correlation within the predefined window**



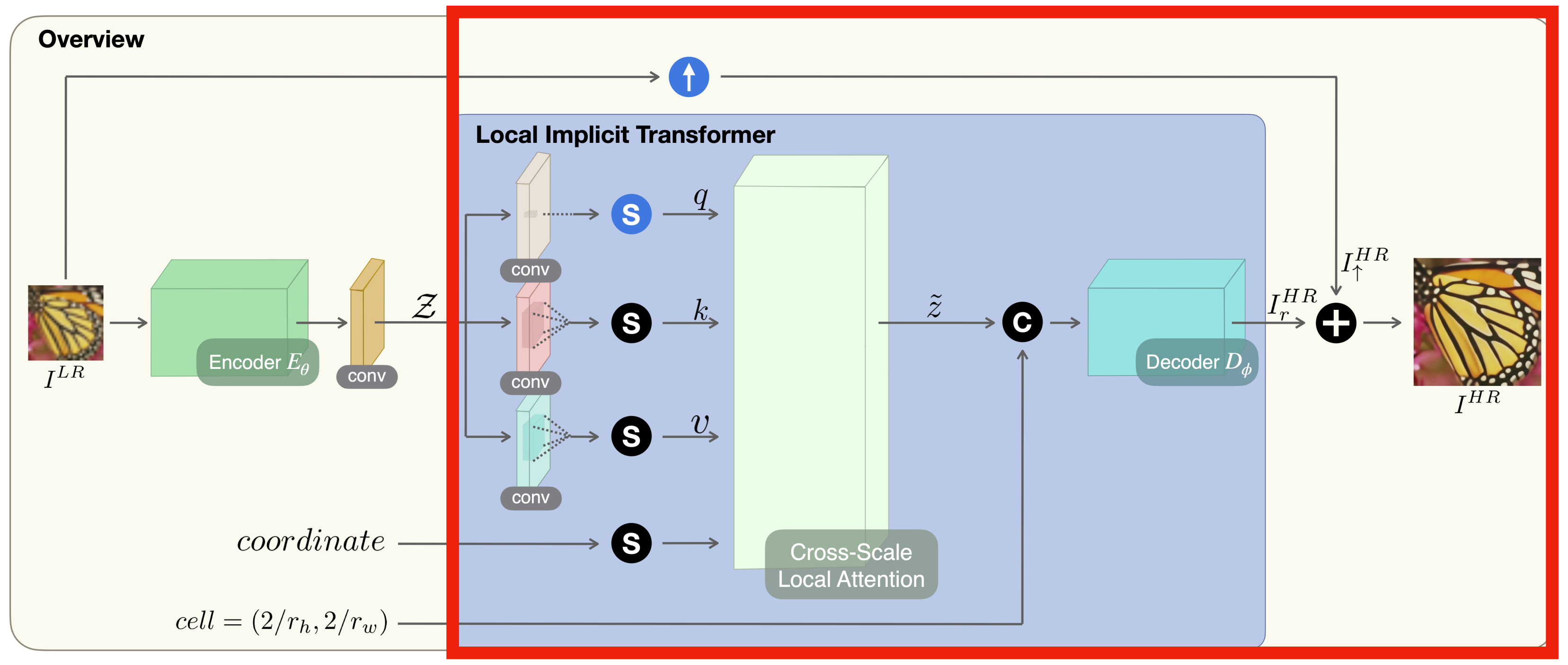
# Methodology

# Methodology

## Architecture



# Methodology



## LIT - Problem formulation

**We applying a local attention mechanism on the latent embedding to take feature correlation and relative distances into consideration at the same time**

- **Goal.** *Learn the residual term*

The queried HR coordinate

$$I^{\mathrm{HR}}(x_q) = I_\uparrow^{\mathrm{HR}}(x_q) + I_{\mathrm{r}}^{\mathrm{HR}}(x_q)$$

- **LIT module configuration**

Cell size (pixel size)

$c = (2/s_h, 2/s_w)$

$s_h$ and $s_w$ indicate height and width of LR image, respectively

$$I_{\mathrm{r}}^{\mathrm{HR}}(x_q) = \mathrm{LIT}(\mathcal{Z}, \delta\mathbf{x}, c)$$

Latent embedding

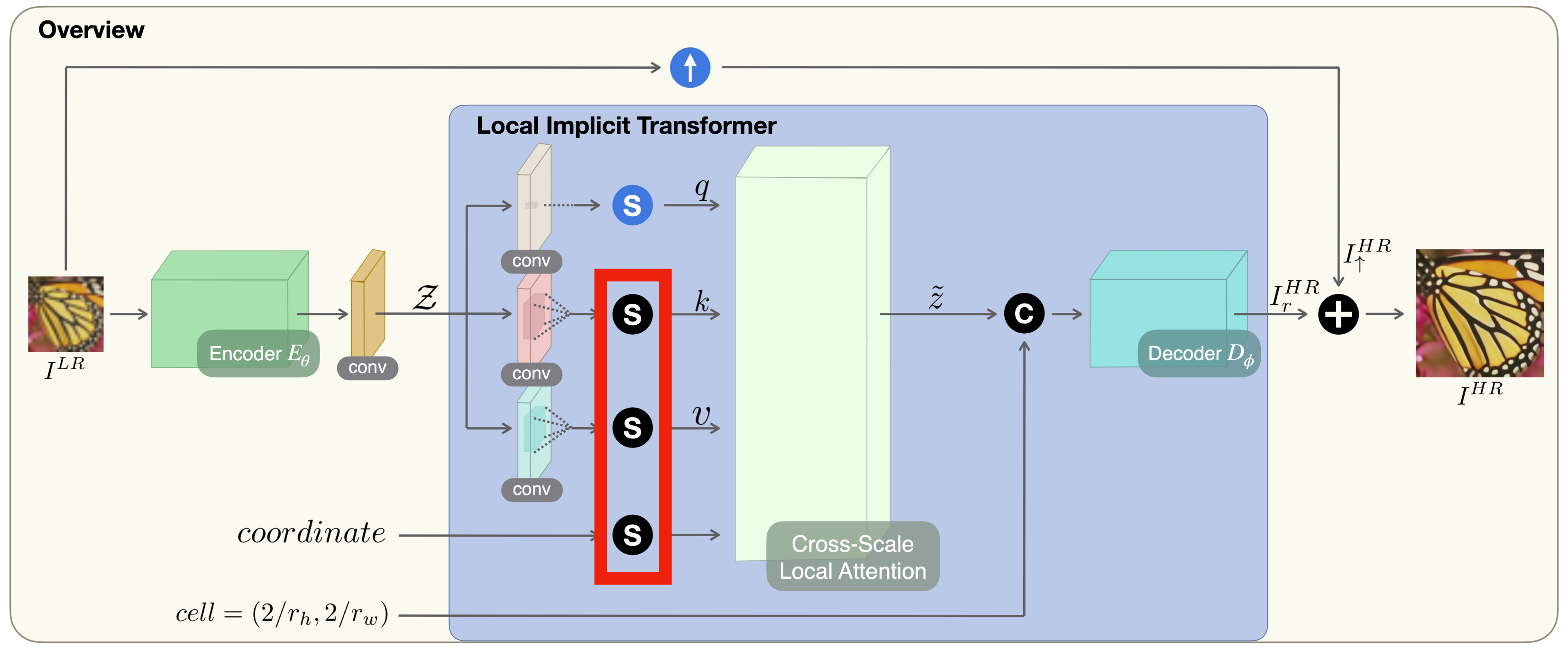
Relative coordinates

$\delta\mathbf{x} = \{x_q - v^*\}$

$v^*$ represents the corresponding nearest LR coordinate

# Methodology

## LIT - Local sampling

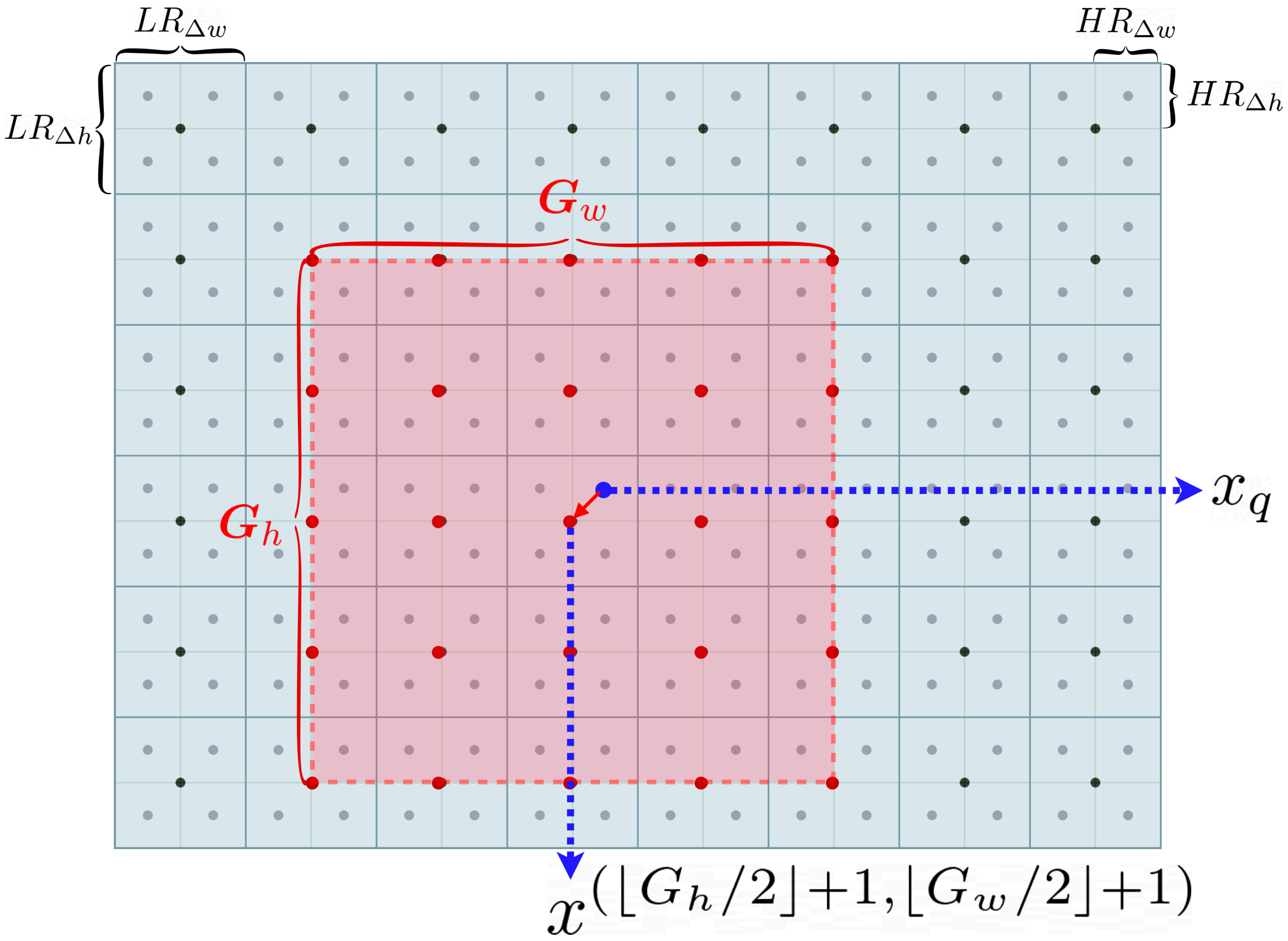


$(LR_{\triangle h}, LR_{\triangle w})$

=> LR pixel size

$(G_h, G_w)$

=> Local grid size

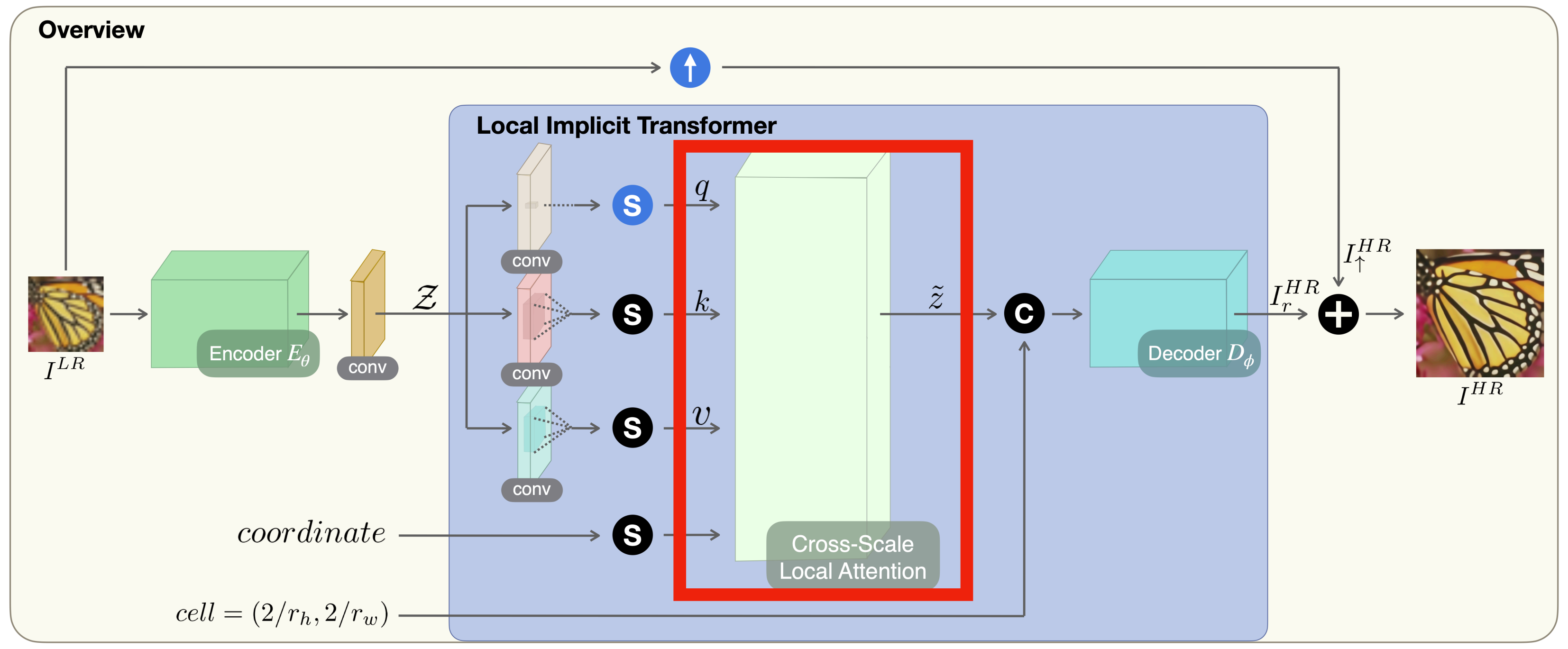


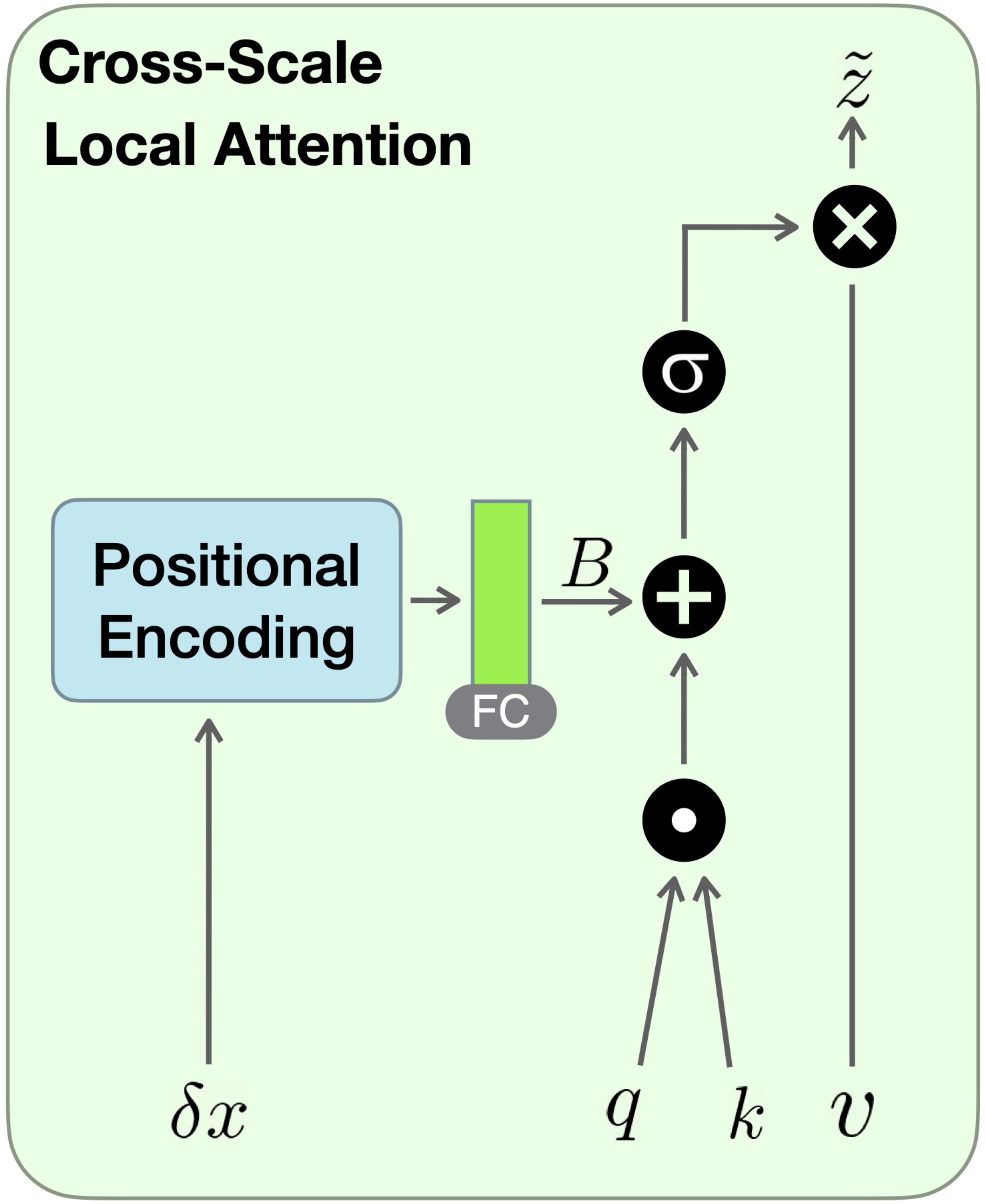
$(HR_{\triangle h}, HR_{\triangle w})$

=> HR pixel size

# Methodology

## LIT - Cross-scale local attention





- **Local attention formulation**

**Feature correlation**

$$\tilde{z} = softmax(\frac{qk^\top}{\sqrt{C}} + B) \times v$$

**Local latent embedding**

**Relative distance**

- **Positional bias term**

$$B = FC(\gamma(\delta \mathbf{x}))$$

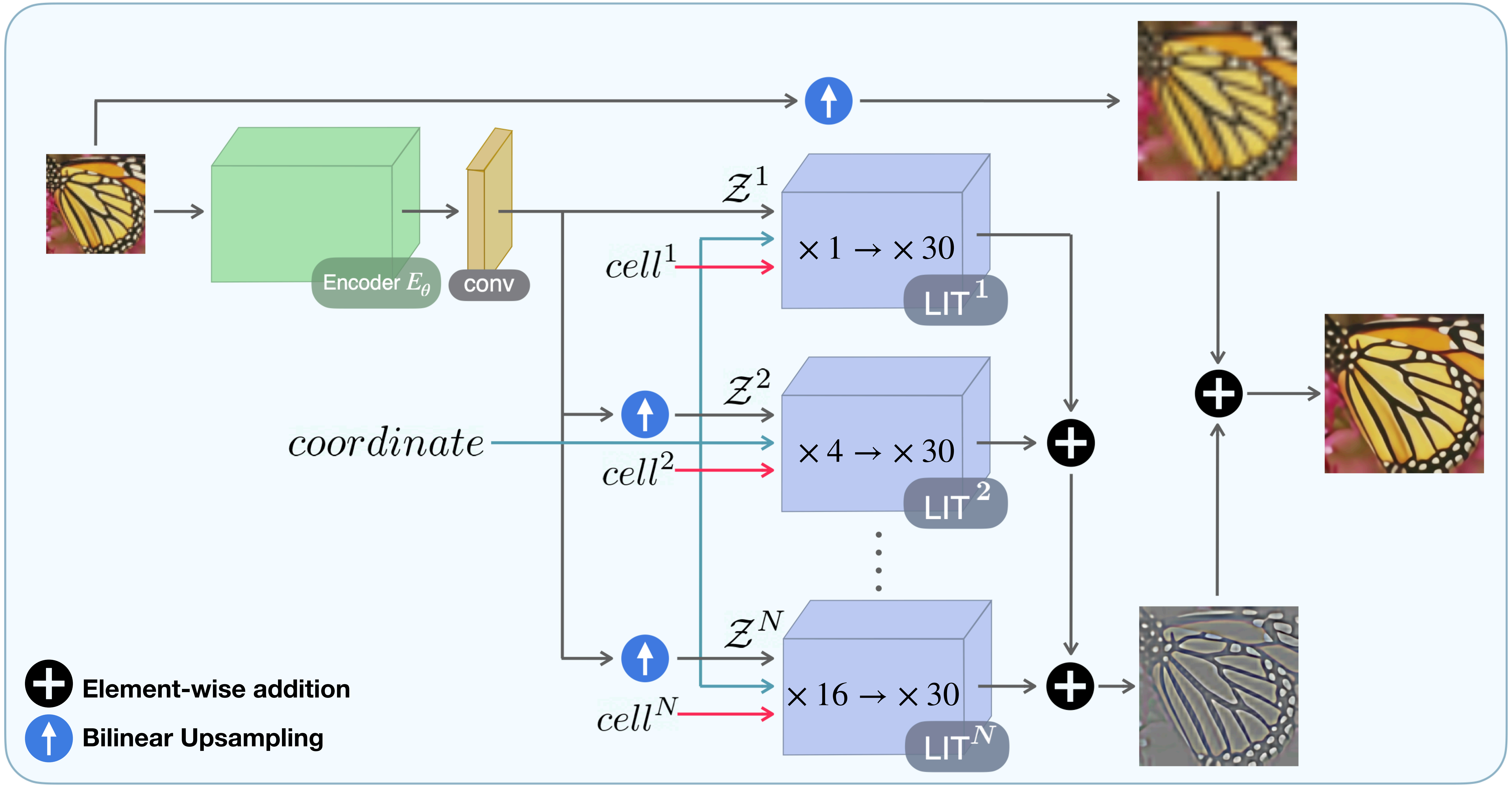
$$\gamma(\delta \mathbf{x}) = [\sin(2^0 \delta \mathbf{x}), \cos(2^0 \delta \mathbf{x}), ..., \sin(2^{L-1} \delta \mathbf{x}), \cos(2^{L-1} \delta \mathbf{x})]$$

⊙ **Inner product**
σ **Softmax**
⊕ **Element-wise addition**
⊗ **Element-wise multiplication**

# Methodology

## Cascaded LIT (CLIT)

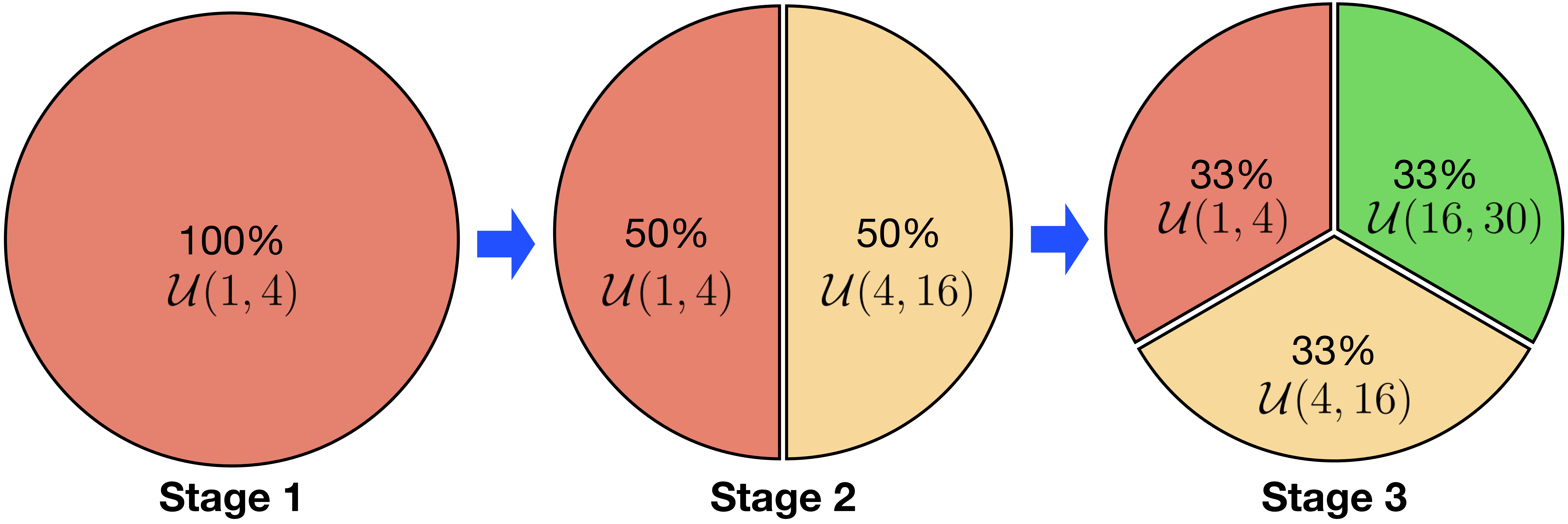
**Extend one-step to multi-step upscaling when dealing with larger upsampling factors (e.g., x30)**



# Methodology

## Cascaded LIT (CLIT) - Cumulative training strategy

**Instead of sampling an upsampling factor uniformly to train the model, CLIT alternatively switches between small and large upsampling factors (e.g., x30)**

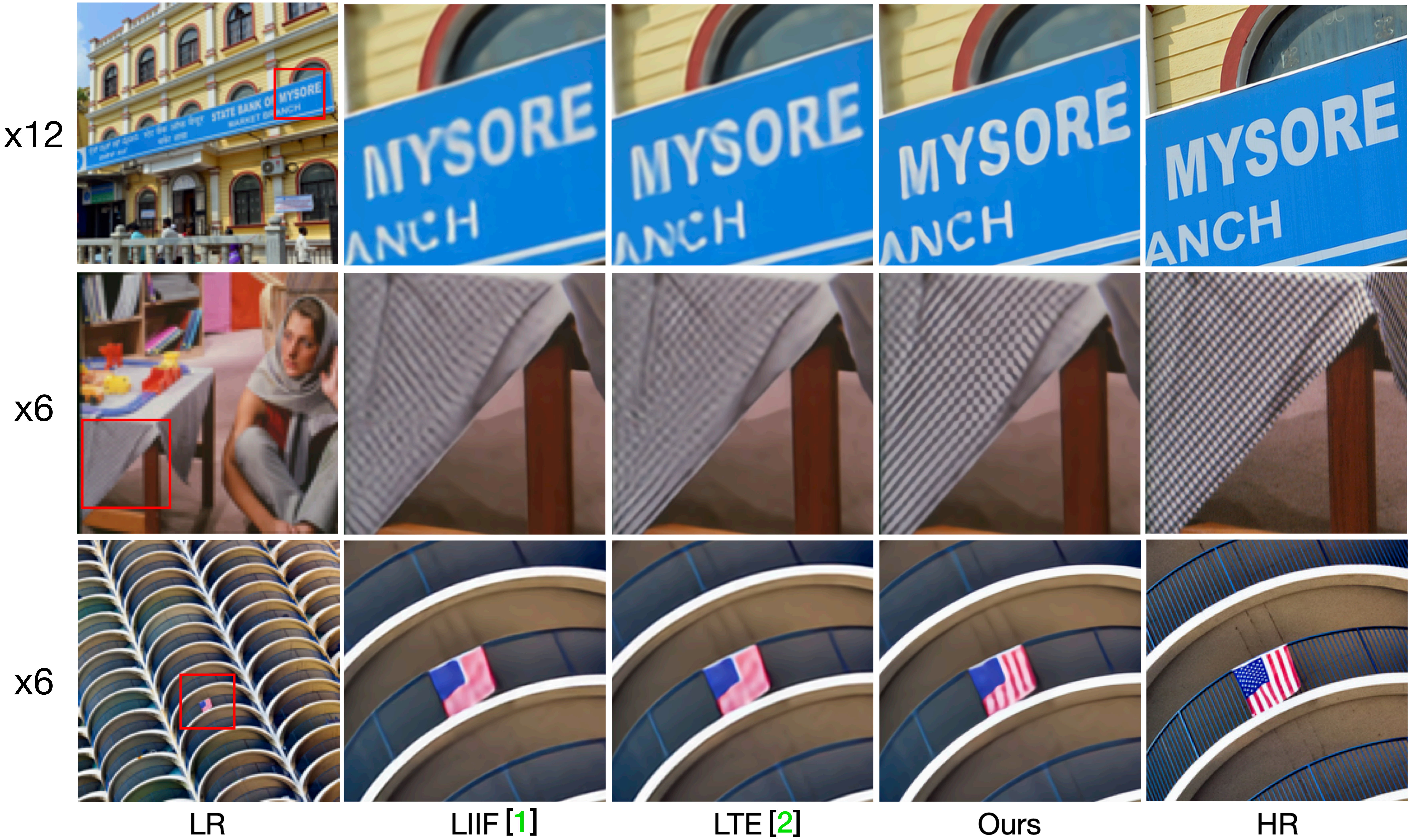


# Experimental Results

# Experimental Results

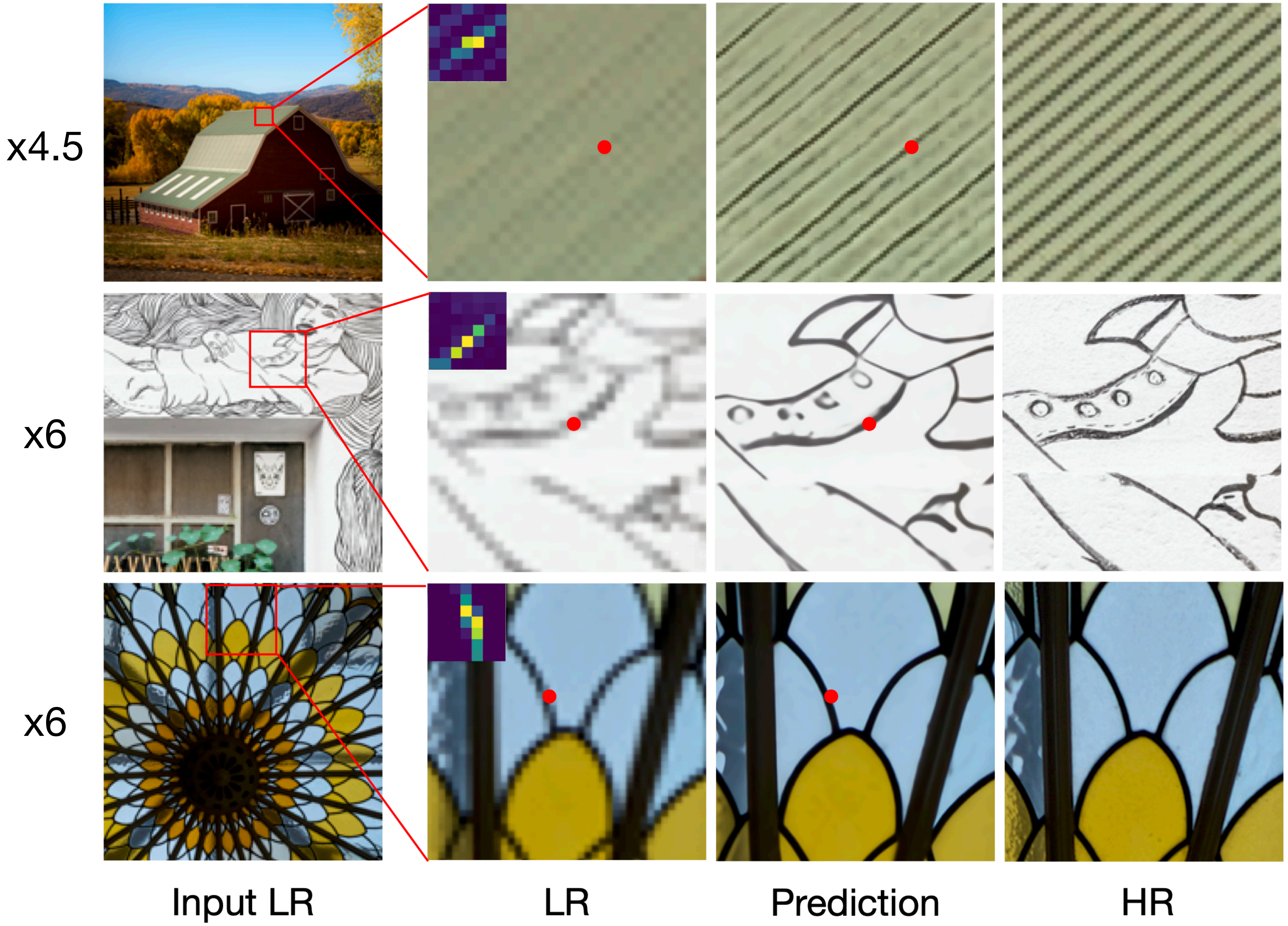
[1] Y. Chen *et al.*, Learning continuous image representation with local implicit image function, CVPR 2021.
[2] J. Lee *et al.*, Local texture estimator for implicit representation function, CVPR 2022.

## Qualitative results - Baseline comparison



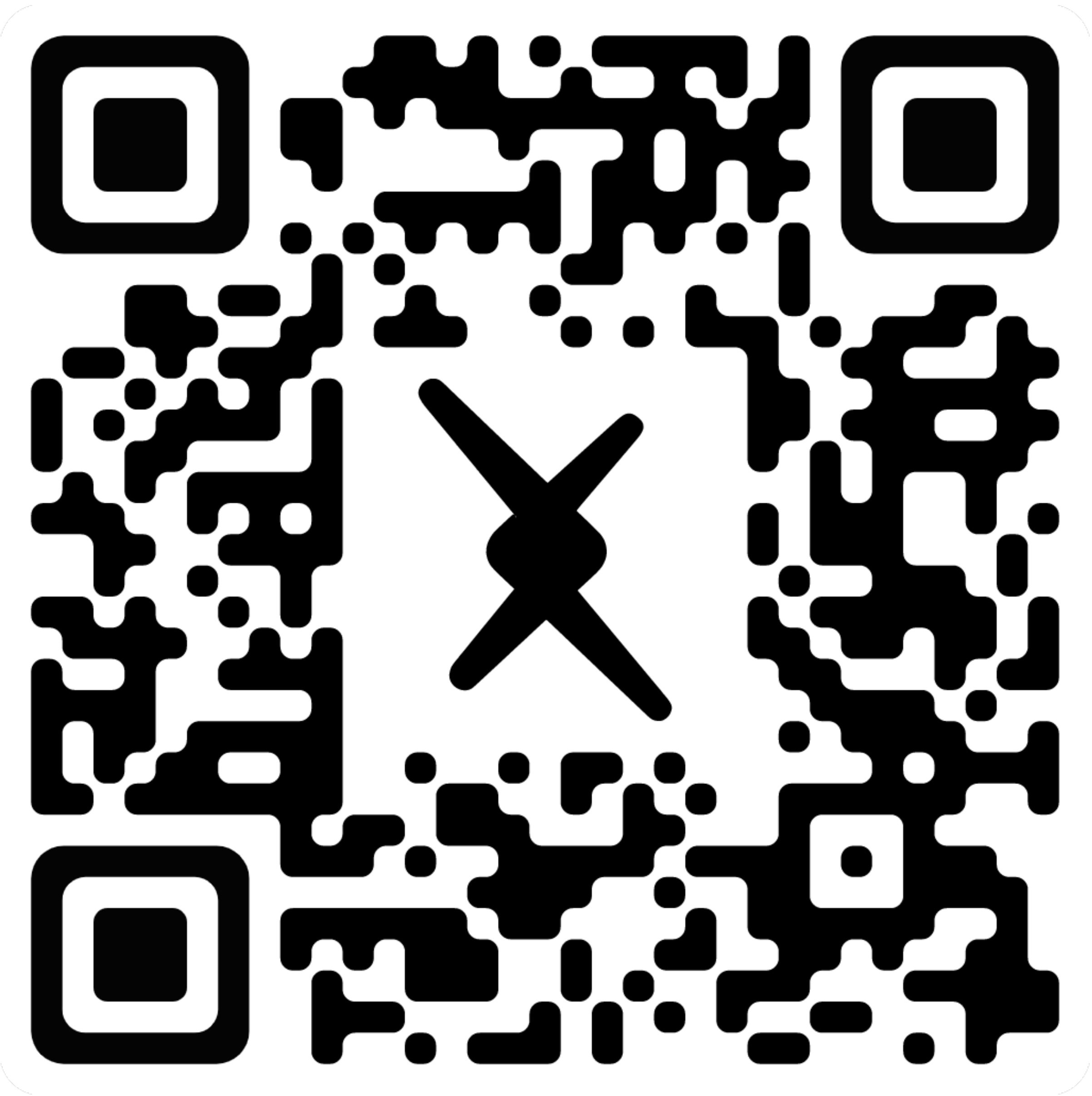
# Experimental Results

## Qualitative results - Attention maps



# Video Demonstration and Quantitative Results

- The first QR code links to another YouTube video, which provides more qualitative results of our work.
- If you are interested in the quantitative results, please refer to our paper, which can be accessed through the second QR code.



## Thank you for your listening

# Cascaded Local Implicit Transformer for Arbitrary-Scale Super-Resolution (CLIT)

**Hao-Wei Chen**[*,1,2], **Yu-Syuan Xu**[*,1,2], Min-Fong Hong[2], Yi-Min Tsai[2], Hsien-Kai Kuo[2], and Chun-Yi Lee[1]

*Equal contribution [1]Elsa Lab, National Tsing Hua University [2]MediaTek Inc.

**THU-AM-170**

JUNE 18-22, 2023 CVPR VANCOUVER, CANADA

elsalab.ai

Question ?

jaroslaw1007@gapp.nthu.edu.tw
yu-syuan.xu@mediatek.com
cylee@cs.nthu.edu.tw