上海科技大学
ShanghaiTech University

Project-1: Deep Burst Super-Resolution

Yujiao Shi
SIST, ShanghaiTech
Spring, 2024

Outline

- Task
- Method & Framework
- Training Objective
- Data Processing
- Project Requirement

Task

Deep Burst Super-Resolution

Goutam Bhat Martin Danelljan Luc Van Gool Radu Timofte
Computer Vision Lab, ETH Zurich, Switzerland

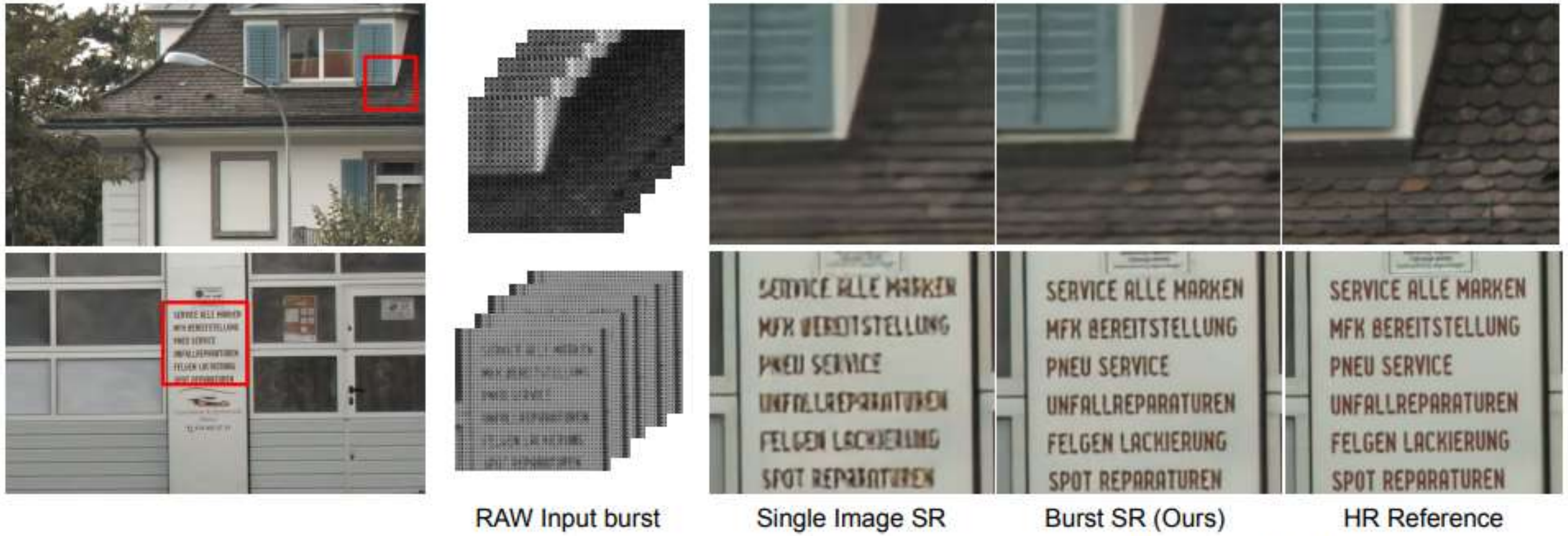
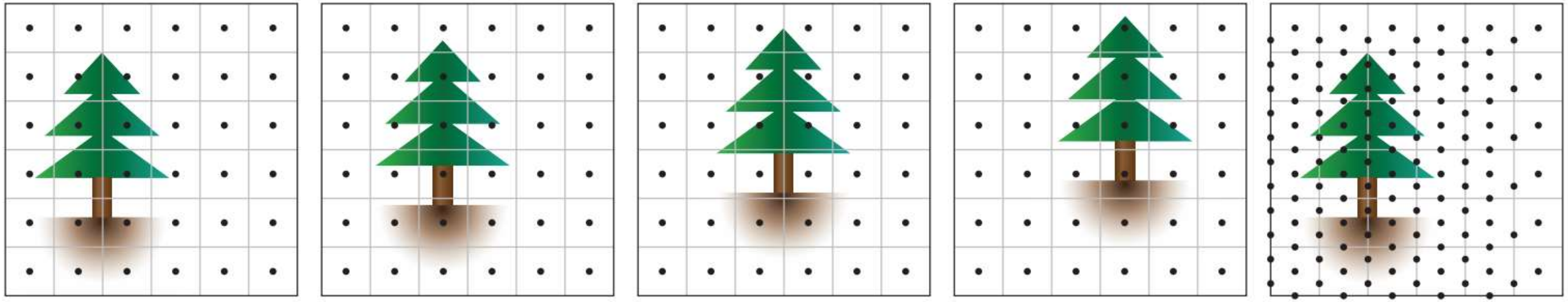


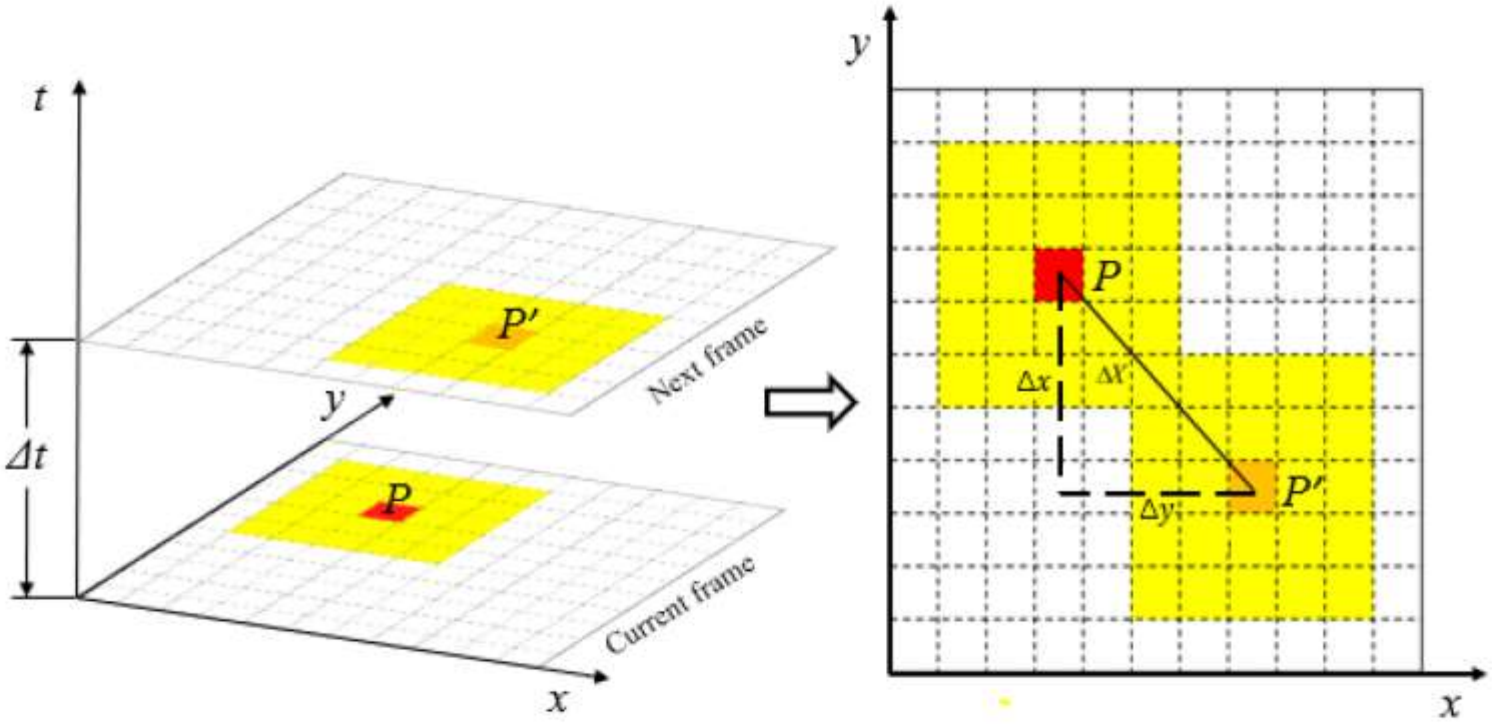
Figure 1. Our network generates a super-resolved RGB image from an input burst consisting of multiple noisy RAW frames. In contrast to the single image baseline, our approach combines information from multiple frames to obtain a more detailed reconstruction of the scene. The results shown are for super-resolution by a factor of 4.

Why “Burst” images?

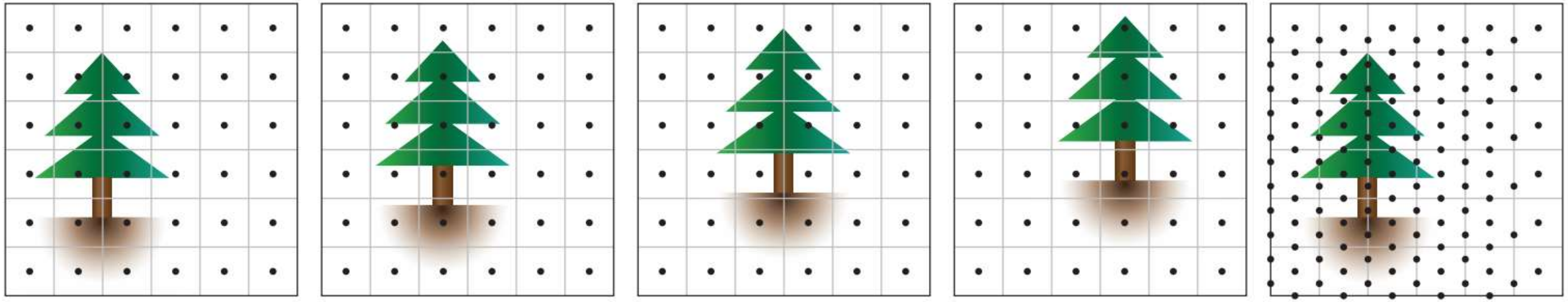


1st frame (base frame)	2nd frame	3rd frame	4th frame	All frames aligned to base frame

What is Optical Flow?



Burst super-resolution

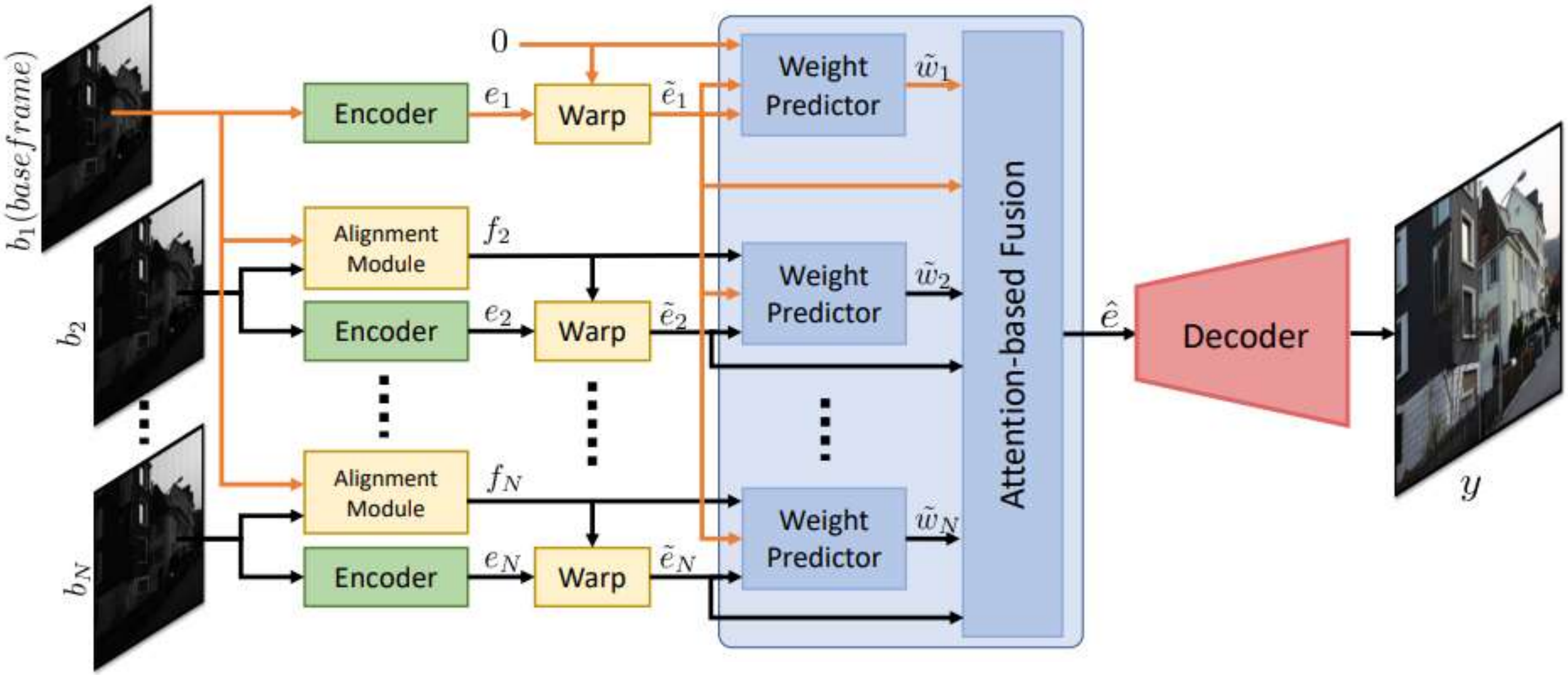


1st frame (base frame) | 2nd frame | 3rd frame | 4th frame | All frames aligned to base frame

$$\hat{e} = \sum_{i=1}^{N} w_i \cdot e_i \,, \quad w_i = \frac{e^{\tilde{w}_i}}{\sum_j e^{\tilde{w}_j}}, \quad \tilde{w}_i = W(\tilde{e}_1, r_i, \hat{f}_i)$$

Method/Framework

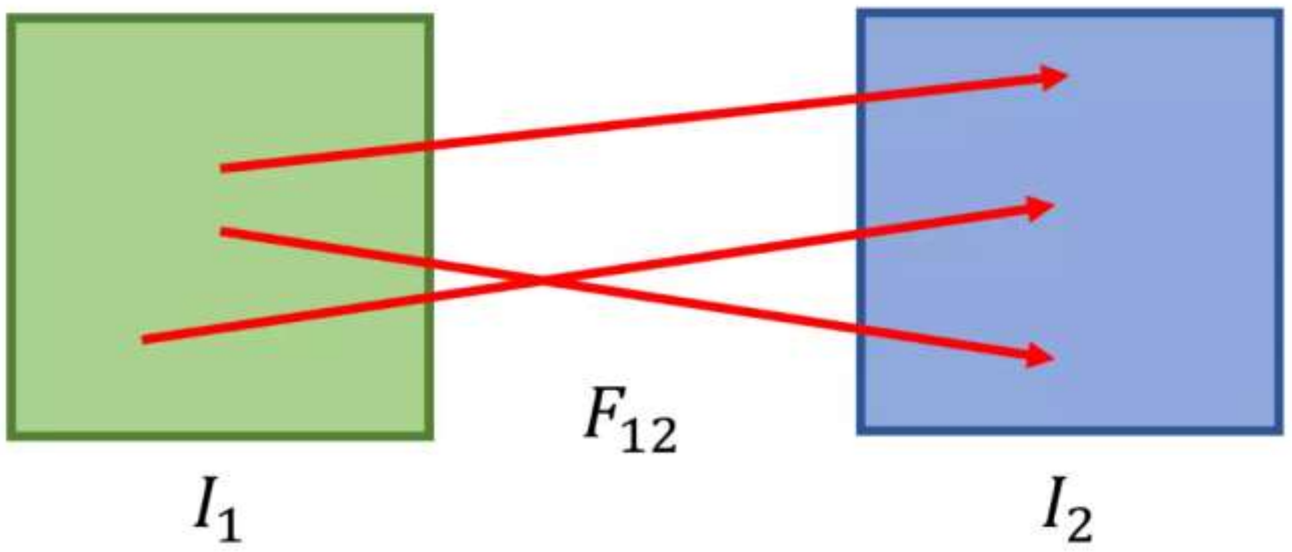
上海科技大学
ShanghaiTech University



$$\hat{e} = \sum_{i=1}^{N} w_i \cdot e_i \,, \quad w_i = \frac{e^{\tilde{w}_i}}{\sum_j e^{\tilde{w}_j}}, \quad \tilde{w}_i = W(\tilde{e}_1, r_i, \hat{f}_i)$$

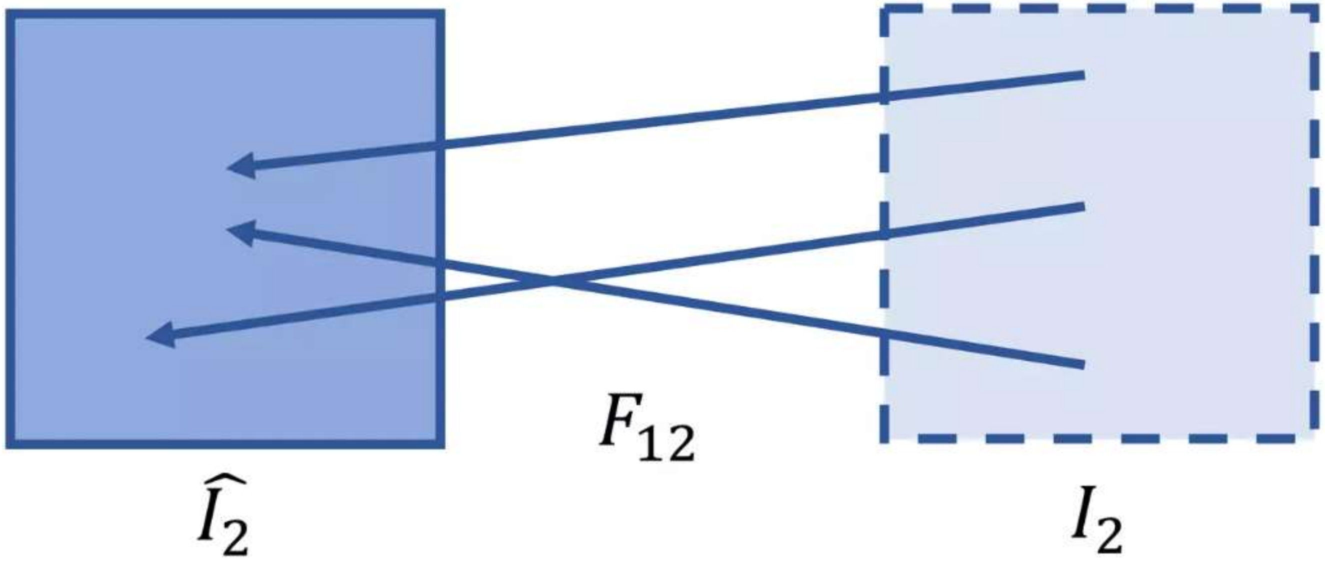
Warping

Warp image I_2 to the frame of I_1



Warping

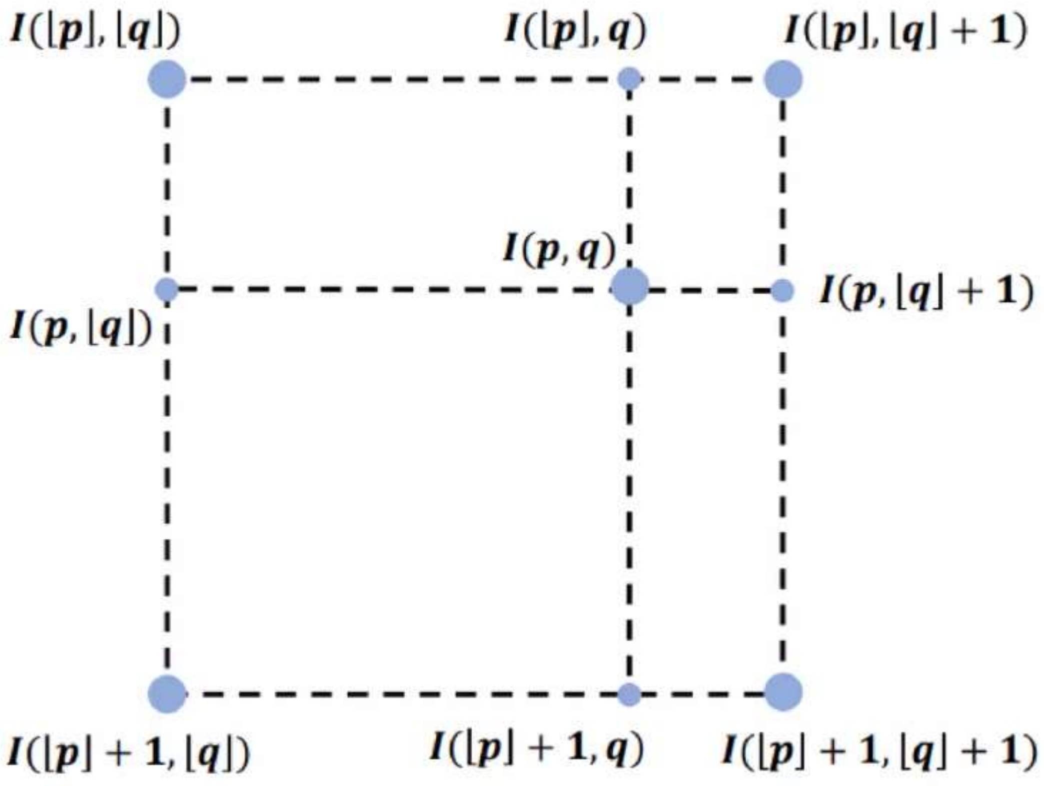
Backward warping



$$\hat{I}_2(x) = I_2(x + F_{12}(x))$$

Warping

上海科技大学 ShanghaiTech University

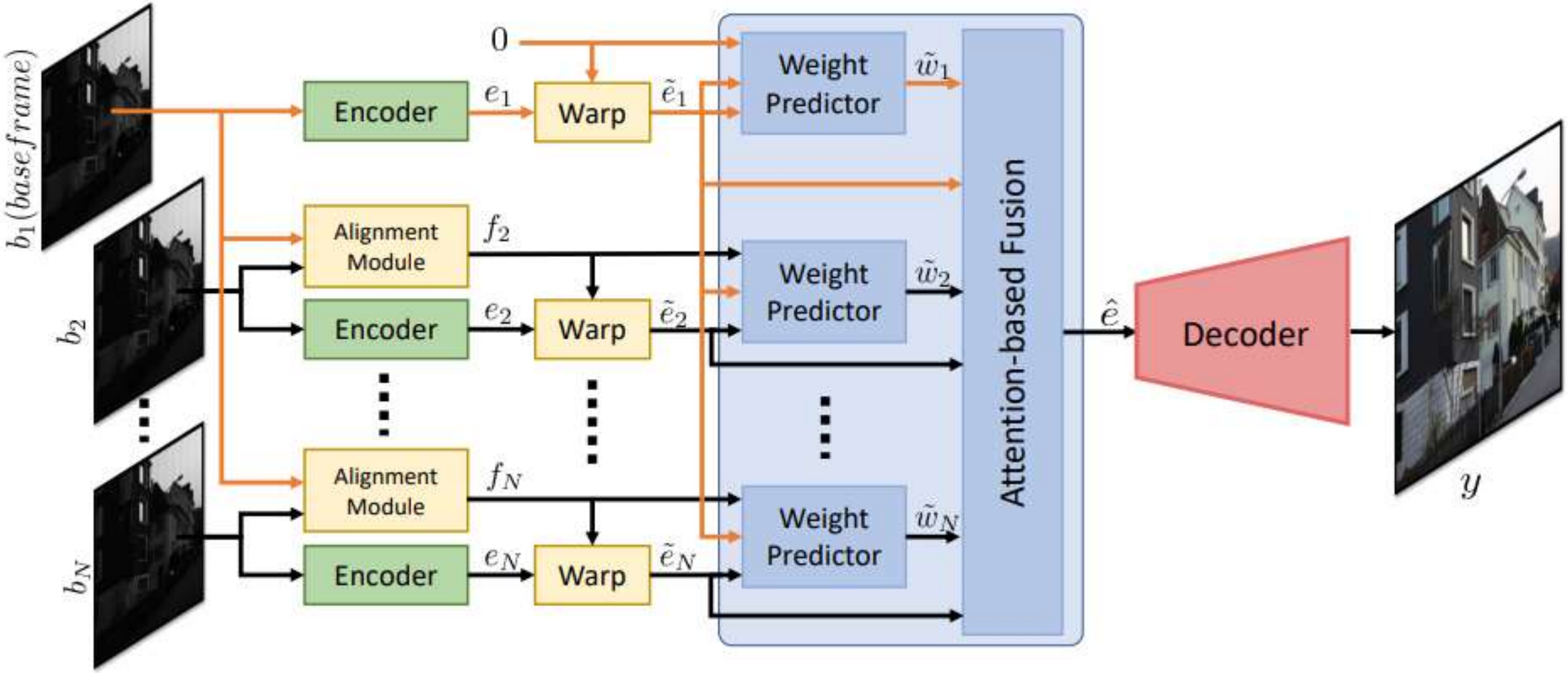


Interpolation

$$
\begin{aligned}
I(p,q) = {} & I(\lfloor p \rfloor, \lfloor q \rfloor) \cdot (1 - p + \lfloor p \rfloor)(1 - q + \lfloor q \rfloor) \\
& + I(\lfloor p \rfloor + 1, \lfloor q \rfloor) \cdot (p - \lfloor p \rfloor)(1 - q + \lfloor q \rfloor) \\
& + I(\lfloor p \rfloor, \lfloor q \rfloor + 1) \cdot (1 - p + \lfloor p \rfloor)(q - \lfloor q \rfloor) \\
& + I(\lfloor p \rfloor + 1, \lfloor q \rfloor + 1) \cdot (p - \lfloor p \rfloor)(q - \lfloor q \rfloor),
\end{aligned}
$$

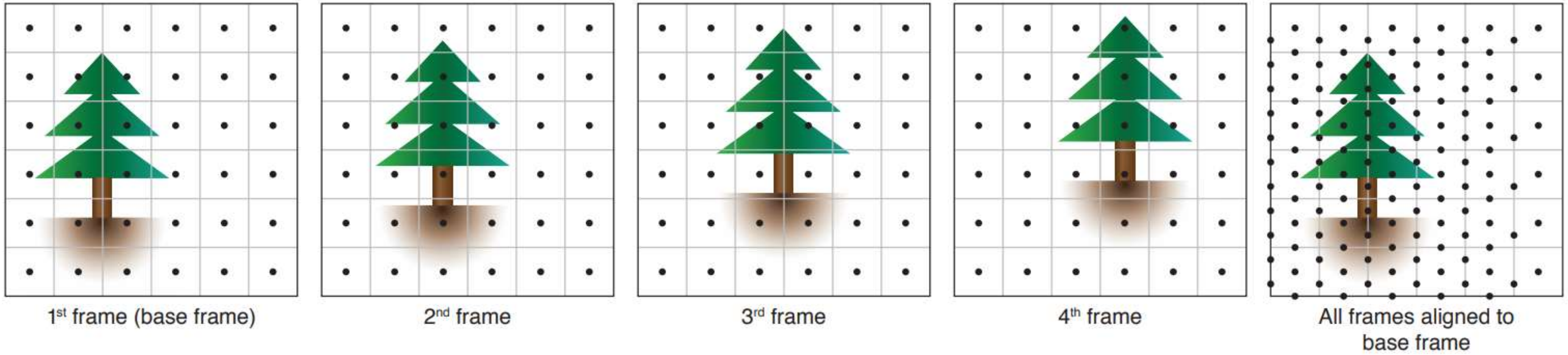
Method/Framework

上海科技大学 ShanghaiTech University



$$\hat{e} = \sum_{i=1}^{N} w_i \cdot e_i\,, \quad w_i = \frac{e^{\tilde{w}_i}}{\sum_j e^{\tilde{w}_j}}, \quad \tilde{w}_i = W(\tilde{e}_1, r_i, \hat{f}_i)$$

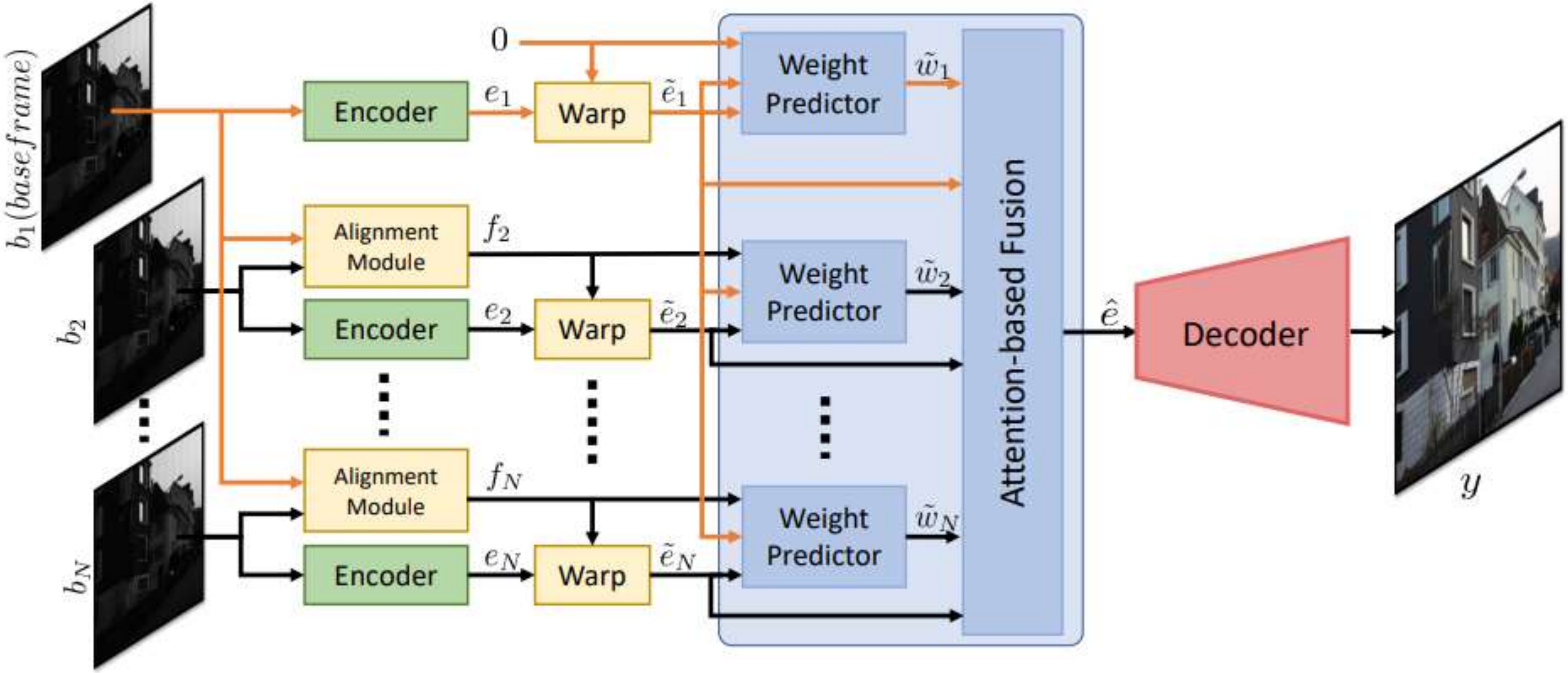
Burst super-resolution



$$\hat{e} = \sum_{i=1}^{N} w_i \cdot e_i \,, \quad w_i = \frac{e^{\tilde{w}_i}}{\sum_j e^{\tilde{w}_j}} \,, \quad \tilde{w}_i = W(\tilde{e}_1, r_i, \hat{f}_i)$$

Method/Framework

上海科技大学
ShanghaiTech University



$$\hat{e} = \sum_{i=1}^{N} w_i \cdot e_i \,, \quad w_i = \frac{e^{\tilde{w}_i}}{\sum_j e^{\tilde{w}_j}}, \quad \tilde{w}_i = W(\tilde{e}_1, r_i, \hat{f}_i)$$

Pixel Shuffle

上海科技大学
ShanghaiTech University

Rearranges elements in a tensor of shape $(*, C \times r^2, H, W)$ to a tensor of shape $(*, C, H \times r, W \times r)$, where r is an upscale factor.

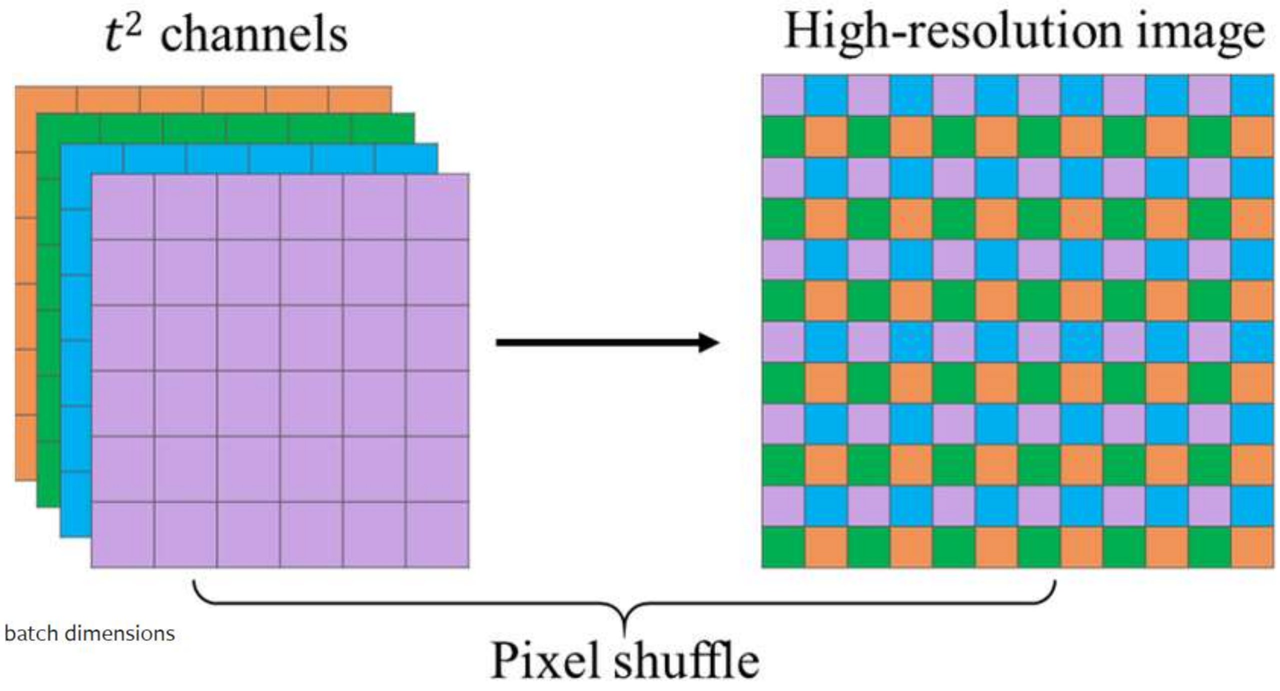
- Input: $(*, C_{in}, H_{in}, W_{in})$, where * is zero or more batch dimensions
- Output: $(*, C_{out}, H_{out}, W_{out})$, where

$$C_{out} = C_{in} \div \text{upscale\_factor}^2$$

$$H_{out} = H_{in} \times \text{upscale\_factor}$$

$$W_{out} = W_{in} \times \text{upscale\_factor}$$

Pixel Shuffle



- Input: $(*, C_{in}, H_{in}, W_{in})$, where * is zero or more batch dimensions
- Output: $(*, C_{out}, H_{out}, W_{out})$, where

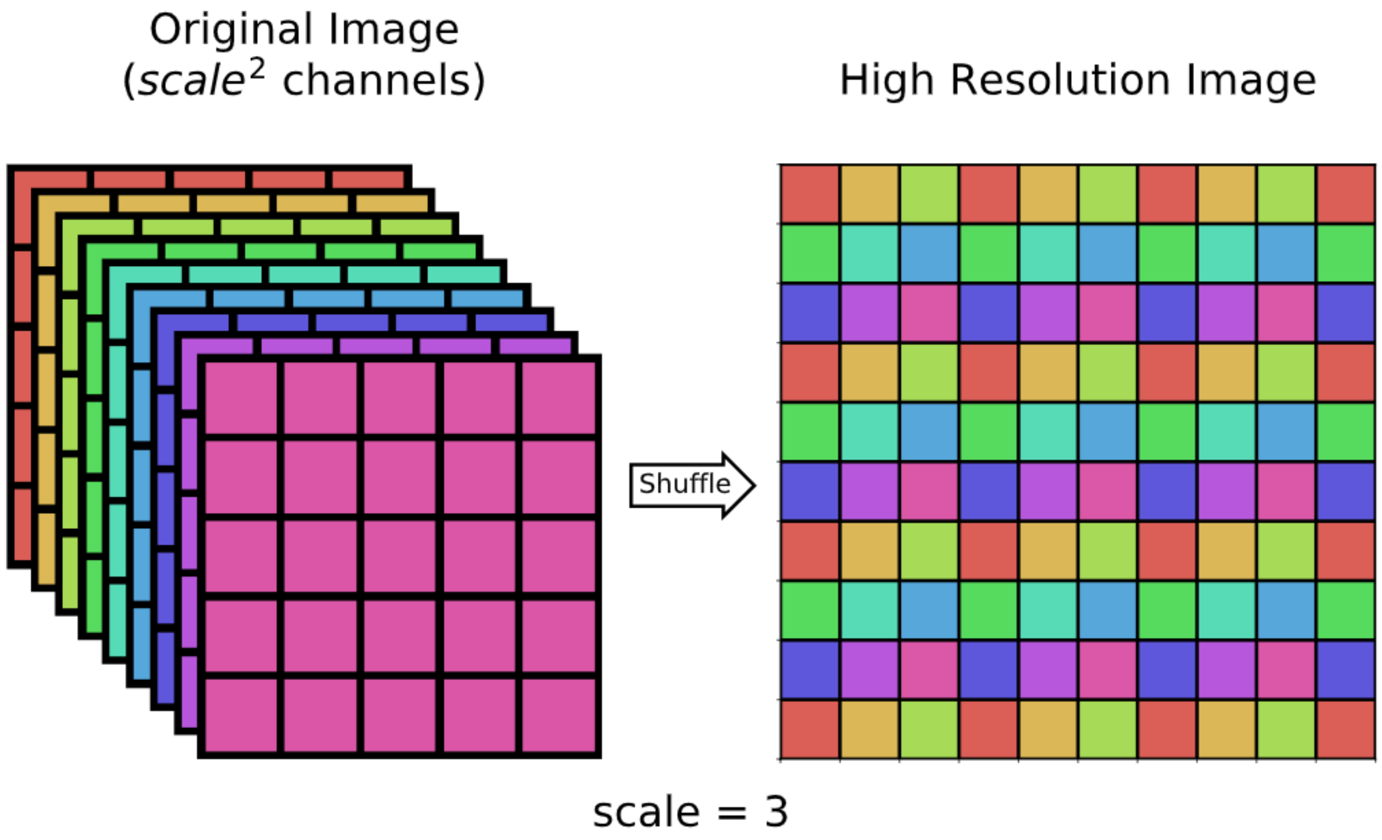
$$C_{out} = C_{in} \div \text{upscale\_factor}^2$$

$$H_{out} = H_{in} \times \text{upscale\_factor}$$

$$W_{out} = W_{in} \times \text{upscale\_factor}$$

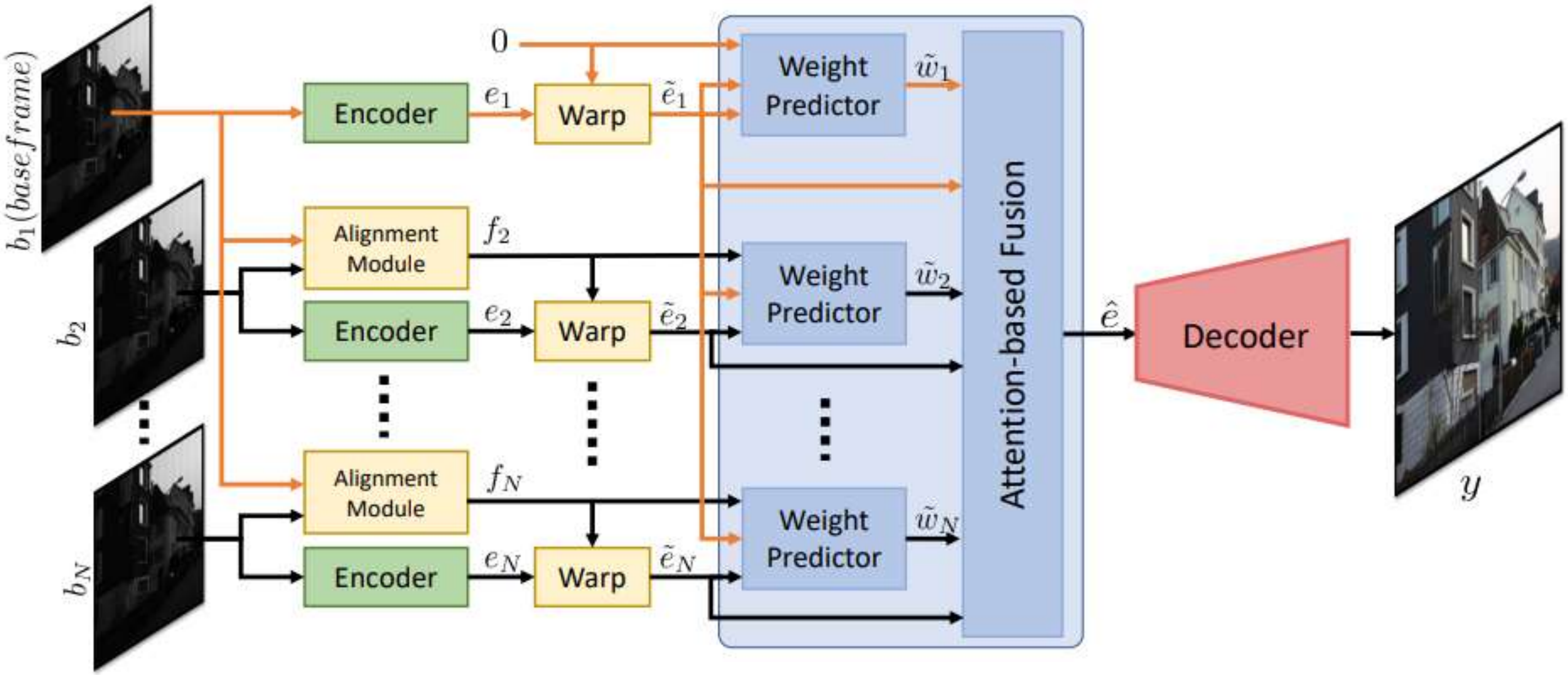
Pixel Shuffle

上海科技大学
ShanghaiTech University



Method/Framework

上海科技大学 ShanghaiTech University



$$\hat{e} = \sum_{i=1}^{N} w_i \cdot e_i\,, \quad w_i = \frac{e^{\tilde{w}_i}}{\sum_j e^{\tilde{w}_j}}, \quad \tilde{w}_i = W(\tilde{e}_1, r_i, \hat{f}_i)$$

Training Objective

- When GT HR image is available:
  - Reconstruction loss (L1 or L2 loss)

- When GT HR image is not available

Pred HR img

$$\ell(y, y_{\mathrm{GT}}) = \sum_n \boxed{m^n} \cdot L_1(\hat{y}^n, y^n_{\mathrm{GT}}), \qquad \hat{y} = \boxed{C}(\phi(\boxed{y}, \boxed{f_{\mathrm{Pred,GT}}}))$$

Binary Mask

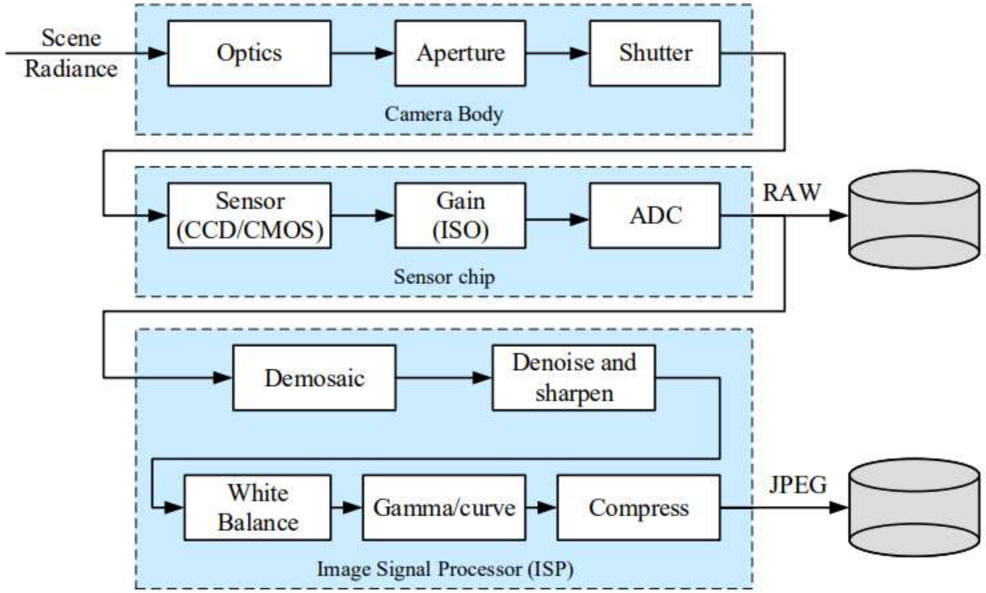
Color Correction Matrix

Optical flow

out image regions which are not aligned correctly. It is set to 0 in regions where the error $R = \left\|\bar{y}_{\mathrm{GT}} - C(\bar{b}_1)\right\|_2$ after color mapping the processed burst image $\bar{b}_1$ is greater than a threshold. Note that the images $\bar{y}_{\mathrm{GT}}$ and $\bar{b}_1$ have lower-

  - It is hard to place the LR & HR camera exactly in the same pose;
  - LR & HR images are from different sensors – color mismatch

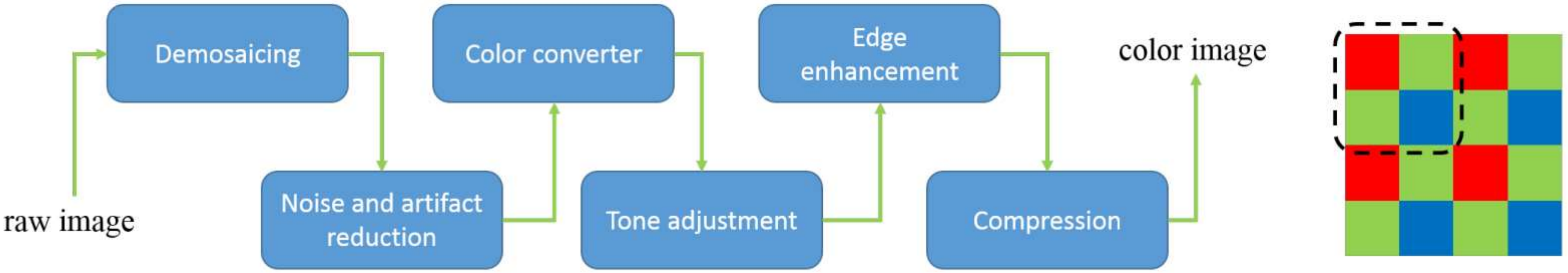
Image Sensing Pipeline



Computer Vision: Algorithms and Applications, 2nd ed.

Image Sensing Pipeline

上海科技大学 ShanghaiTech University

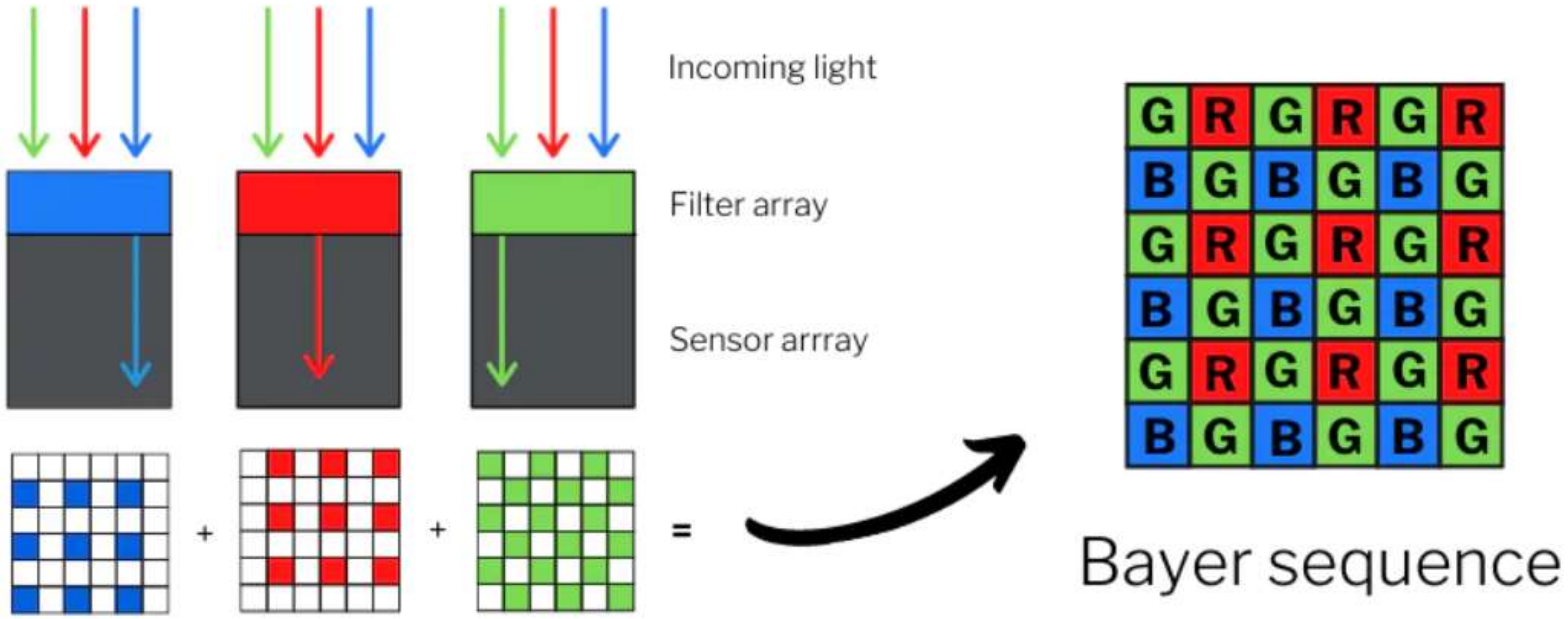


(a) ISP pipeline

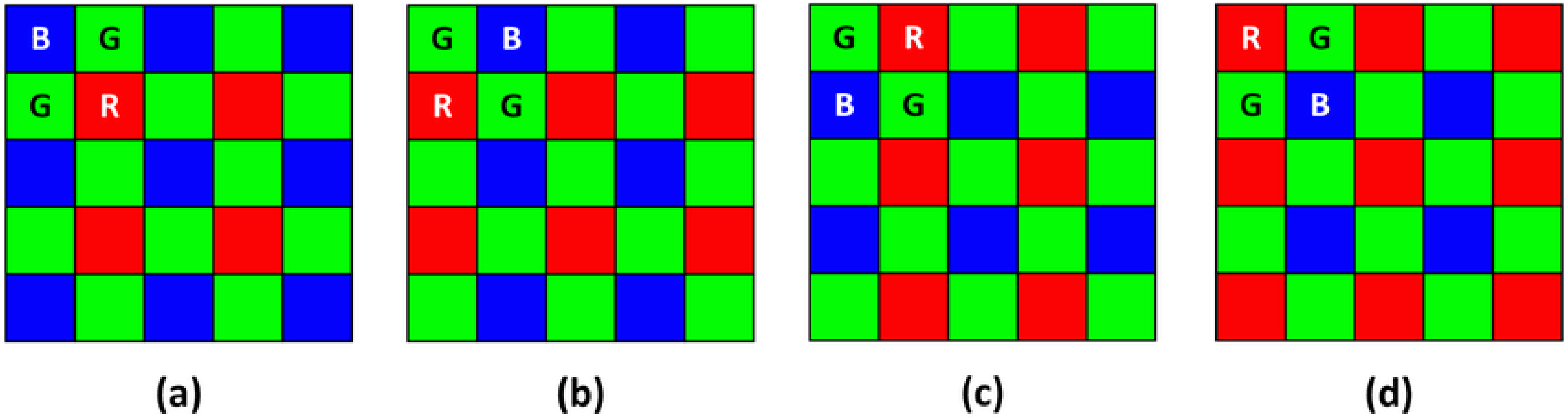
(b) Bayer pattern

Figure from Xu et al., CVPR 2019

Bayer Pattern

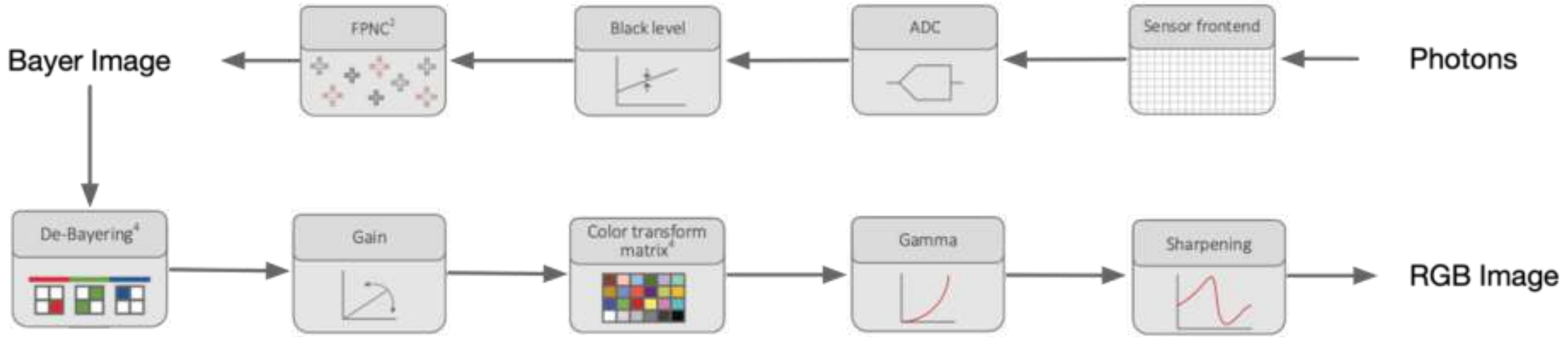


Different Bayer Patterns

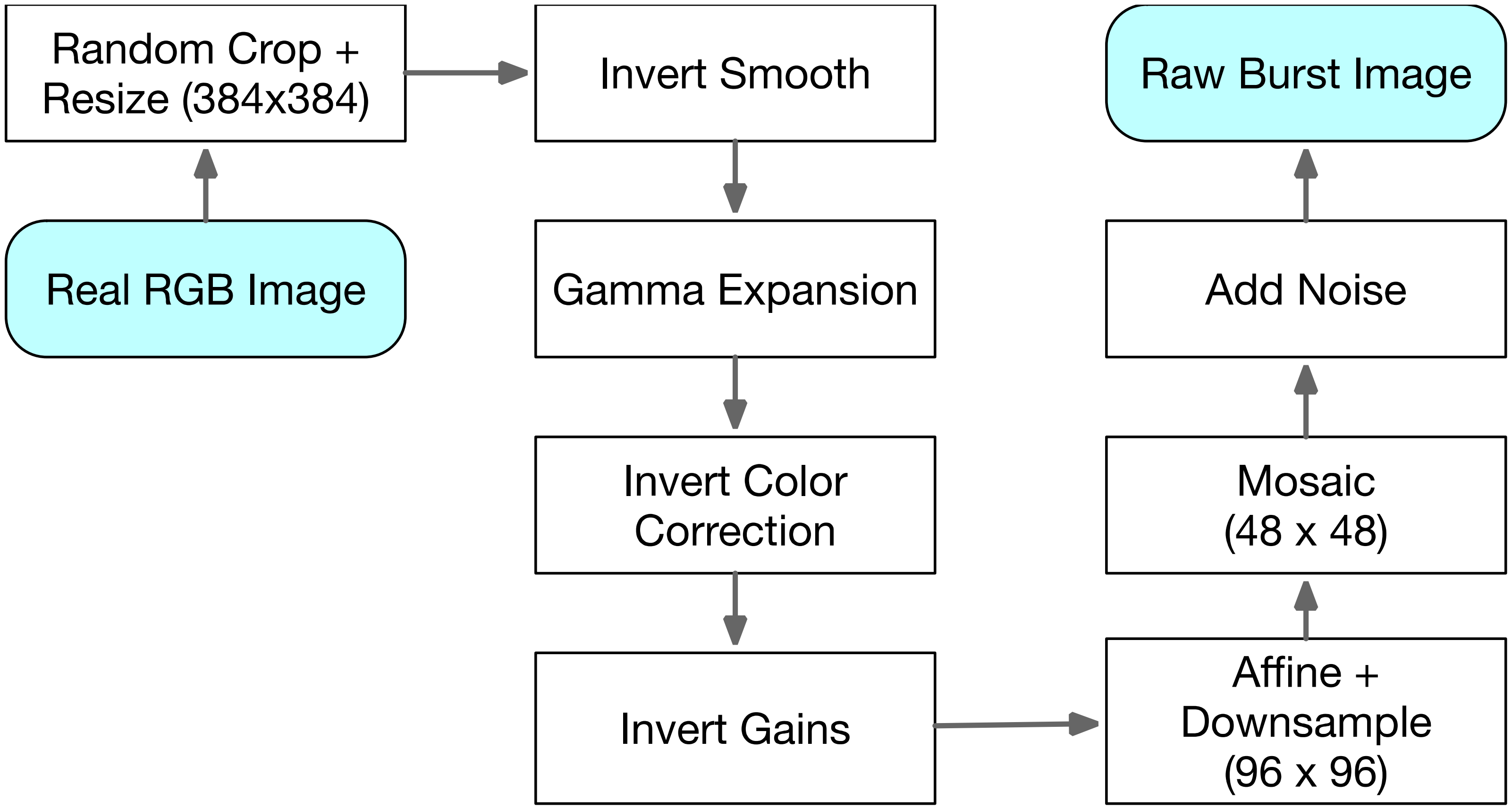


BGGR GBRG GRBG RGGB

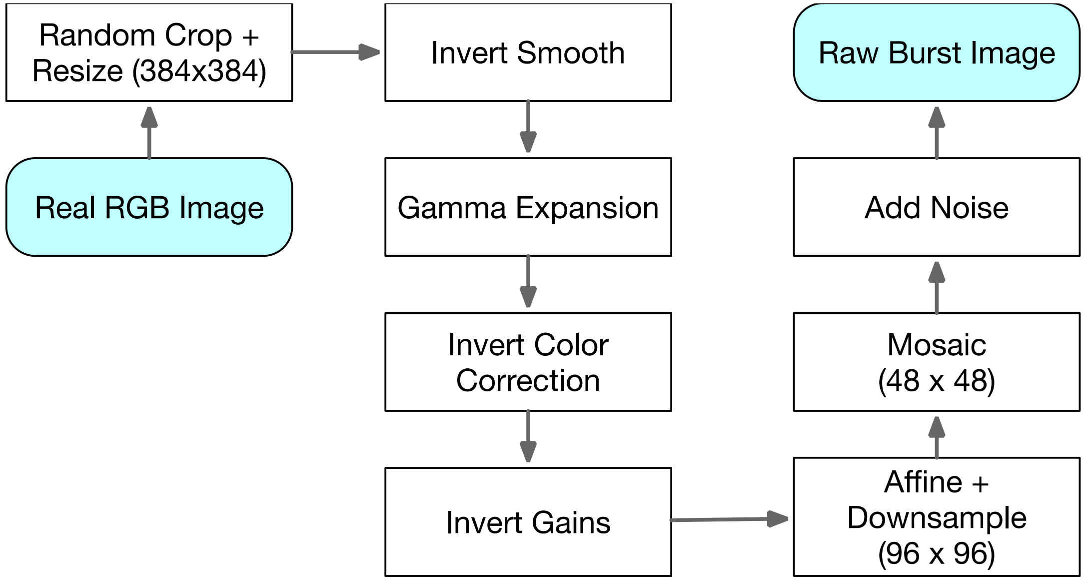
Image Sensing Pipeline



Synthetic Data Creation in DBSR (CVPR 2021)



Synthetic Data Creation in DBSR (CVPR 2021)



Questions:

1. What is the input/output size?
2. Which operation in the model implement the up sampling?
3. What is the up sampling scale?
4. Does the network allow free scales of up sampling?

Project Requirement - Basics

- Testing on Real Data
  - Take an HR image by your own phone;
  - Create LR burst images similar to the synthetic data shown in the paper;
  - Super-resolve the burst images;
  - (Change the Bayer pattern of LR images and super-resolve them again)
  - Compare the similarity between the predicted HR and GT HR images;
  - Capture and evaluate 20 images (repeat the above steps), including both indoor and outdoor scenes; compare and analyze the performance both quantitatively and qualitatively.

- Application to Downstream Tasks
  - Choose a downstream task (semantic segmentation or object detection):
  - apply your LR & HR images to this downstream task;
  - Record and compare the performance difference.

Project Requirement - Advanced

上海科技大学 ShanghaiTech University

- Collect training data by your own phones/cameras:
  - Fine-tune the trained model;
  - Record and compare the performance difference qualitatively and quantitatively.

- Others

References

- Bhat, Goutam, et al. "Deep burst super-resolution." *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2021.
- https://github.com/goutamgmb/deep-burst-sr
- Bhat, Goutam, et al. "Deep reparametrization of multi-frame super-resolution and denoising." Proceedings of the IEEE/CVF international conference on computer vision. 2021.
- Wronski, Bartlomiej, et al. "Handheld multi-frame super-resolution." ACM Transactions on Graphics (ToG) 38.4 (2019): 1-18.
- Xu, Xiangyu, Yongrui Ma, and Wenxiu Sun. "Towards real scene super-resolution with raw images." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019.
- …

Next

上海科技大学
ShanghaiTech University

- Plz start finding your group partners and making preparation for Project-3;

- Every student is required to make presentations during the classes; otherwise your class participation score will be 0 (out of 10).
  - Plz update Tutor Jiawei Yang (yangjw12023@shanghaitech.edu.cn) If you would like to share your super-resolution results on 14th Oct.

- Next – Tutorial on Optical Flow

- Project-2: Camera Localization