ASK 2024 논문집

Annual Symposium of KIPS 2024

신진학자 워크숍

Adaptive Deep Image Signal Processor for Practical Applications

김희원 교수
(숭실대학교)

Adaptive Deep Image Signal Processor *for* Practical Applications

ASK 2024

Heewon Kim

Reality Lab.

Soongsil University

https://reality.ssu.ac.kr

Reality Lab
Soongsil University

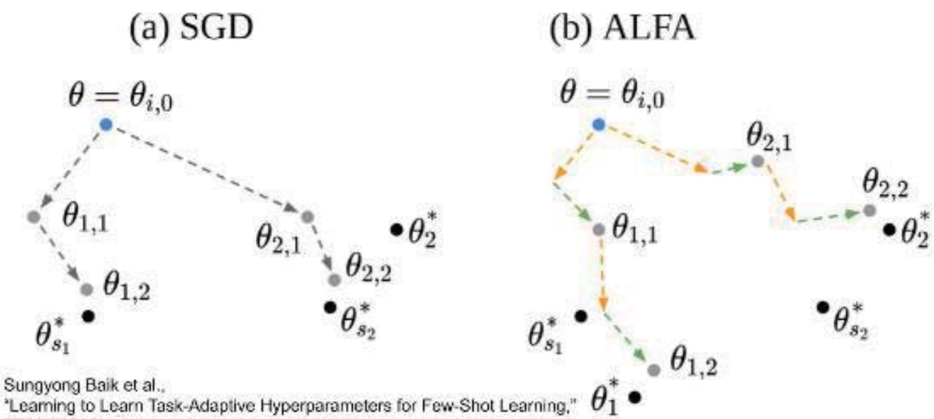
Home · News · Members · Publication · Courses

Welcome to the SSU Reality Lab!

Turning Ideas into Reality

Reality Lab was founded at SSU in 2023. Under the supervision of Prof. Heewon Kim, various topics in virtual and augmented reality have been researched and developed. We mainly focus on extracting and recognizing the information from digital images for algorithms to have artificial intelligence close to humans. Recently, our main interests include: Deep Learning, 2D/3D Vision, Image Restoration, Language Model, AI for Sports, AI in Medicine, AI for Astrophysics, and Media Art.

Recent Publications



Sungyong Baik et al.,
"Learning to Learn Task-Adaptive Hyperparameters for Few-Shot Learning,"
TPAMI 2023

Outline

Introduction

Image Signal Processor: Goal / Challenges / Main Idea (Adaptation)

Proposed Method

Adaptive Data Synthesis / Adaptive Neural Architecture Search / Adaptive ISP Parameter Estimation

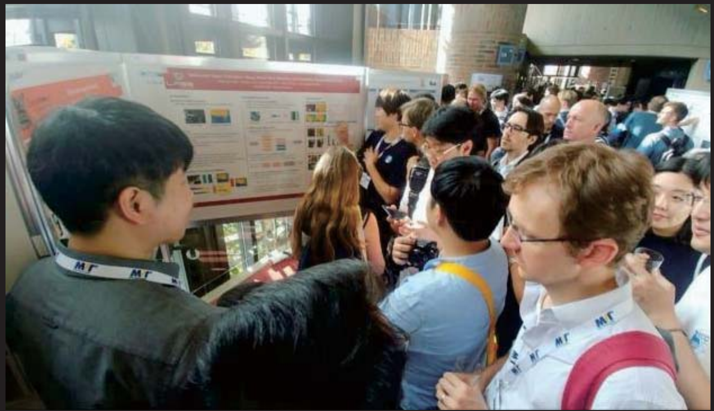
Conclusion

Motivation of Image Signal Processor

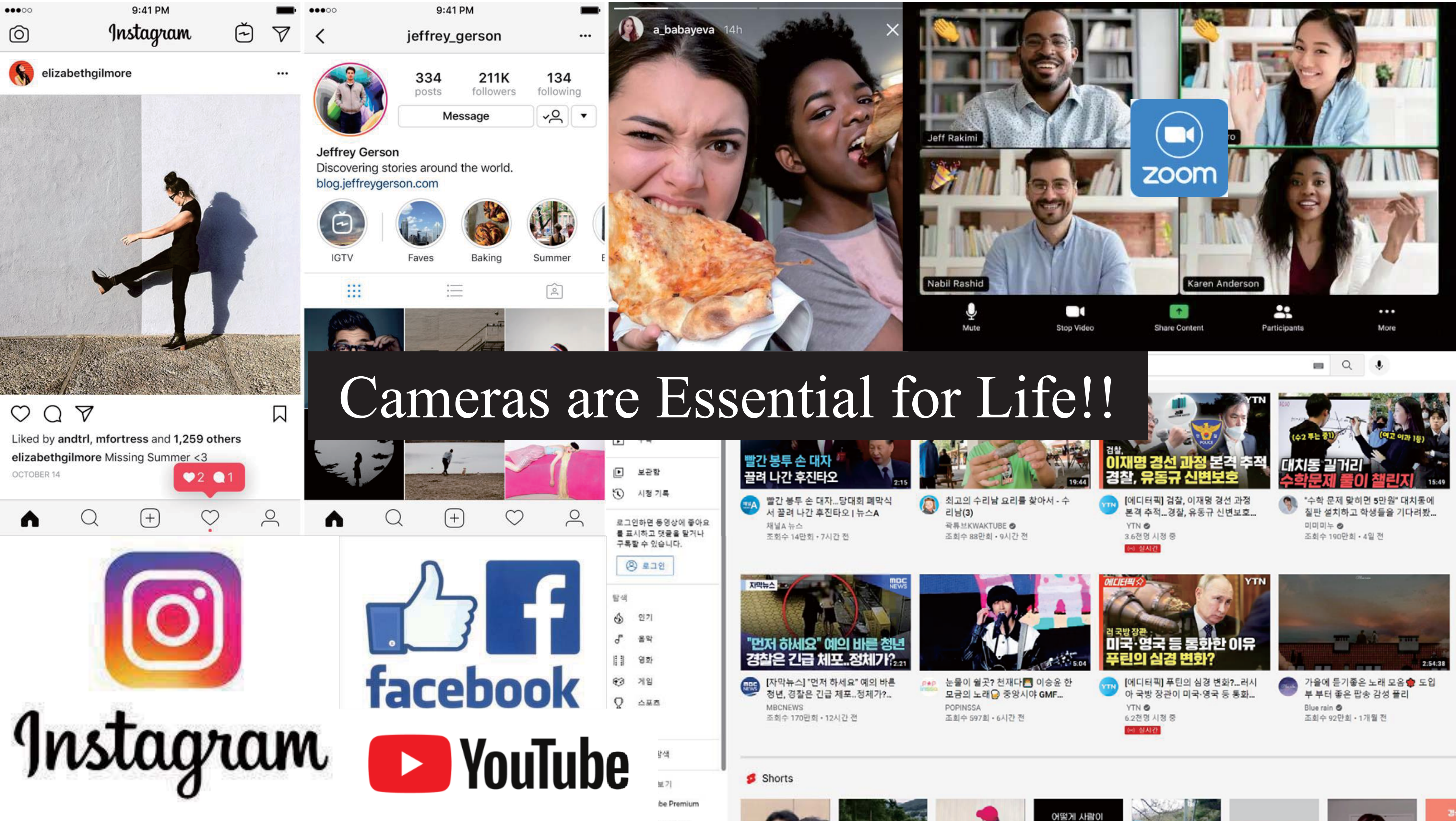
- Taking photos: Various purposes!





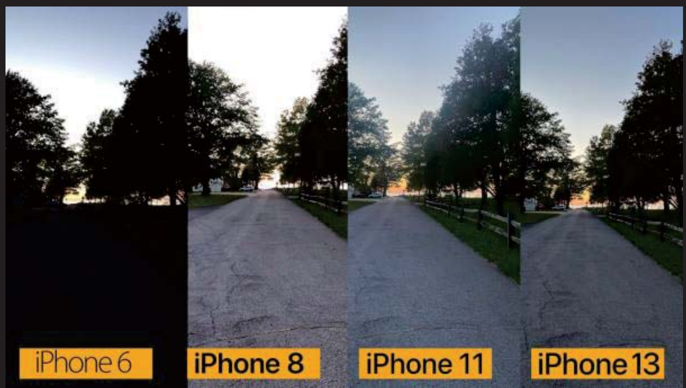



Cameras are Essential for Life!!



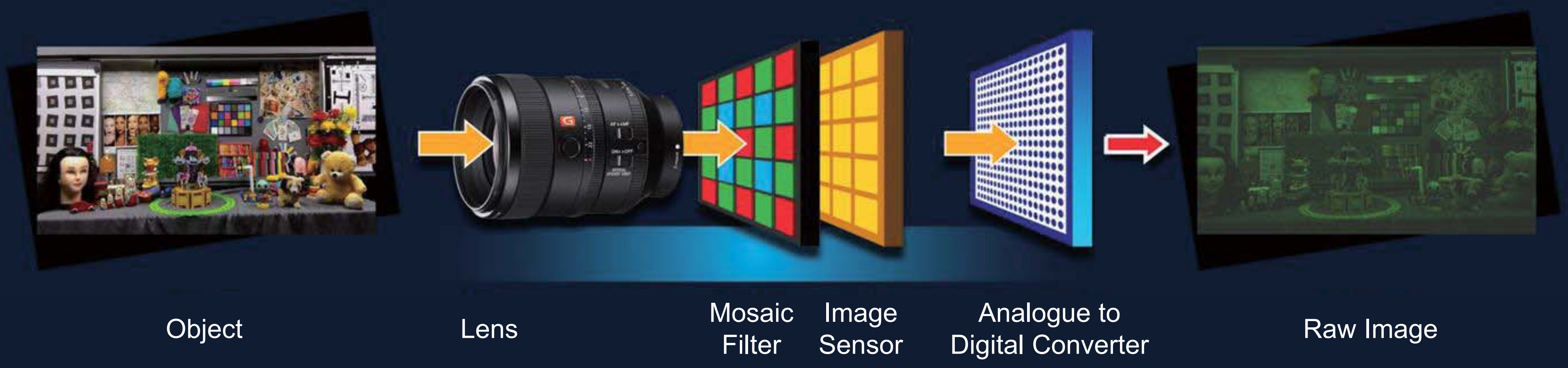
Goal of Image Signal Processor

- The importance of cameras in the smartphone war



Goal of Image Signal Processor

- Digital camera converts lights into digital signals as raw images



Goal of Image Signal Processor

- Digital camera converts lights into digital signals as raw images
- Image Signal Processor (ISP) transforms the raw images to visually pleasing RGB images

Image Signal Processor (ISP)

Raw Image

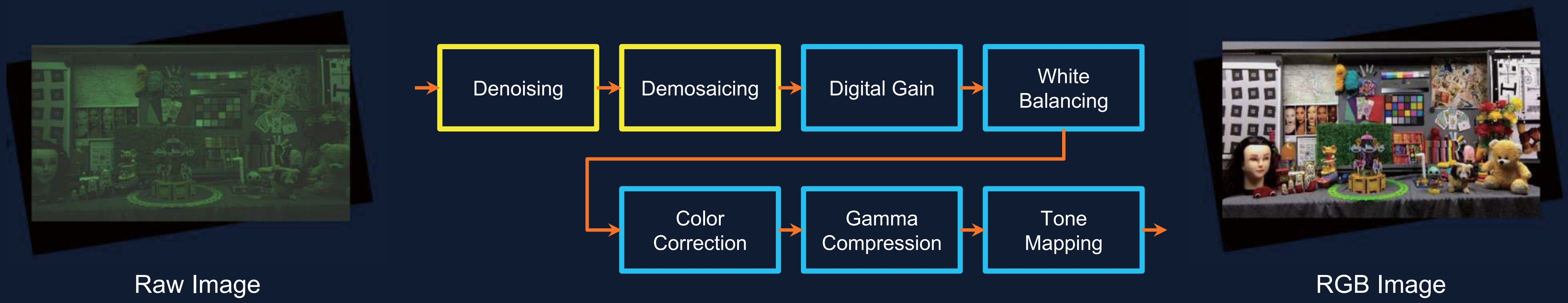
Denoising → Demosaicing → Digital Gain → White Balancing → Color Correction → Gamma Compression → Tone Mapping →

RGB Image

Goal of Image Signal Processor

- Digital camera converts lights into digital signals as raw images
- Image Signal Processor (ISP) transforms the raw images to visually pleasing RGB images
- An ISP performs *image restoration* and *image enhancement*

Image Signal Processor (ISP)



Goal of Image Signal Processor

- Image Restoration
  - Goal: Estimating the original/clean image from a *corrupted* image
  - *Corruption*: Noise, blur, compression, etc

Real scene

RGB Image

Goal of Image Signal Processor

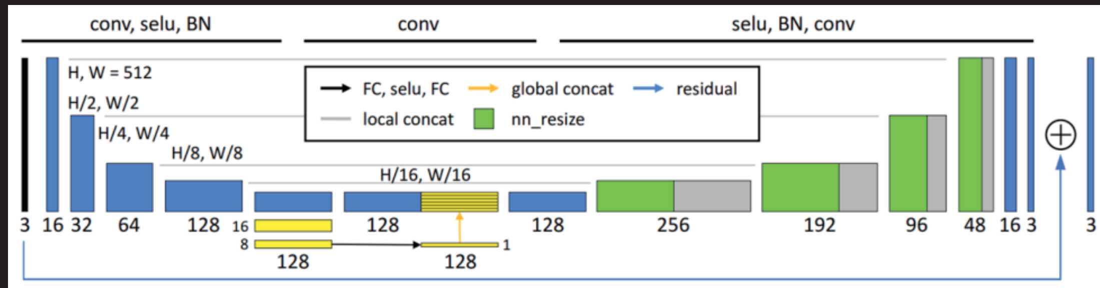
- Image Enhancement
  - Goal: adjusting tone, color, contrast, brightness, and more for *looking better* images
  - *Looking better* (or high-quality) images: Images retouched by trained photographers (MIT-Adobe FiveK, CVPR 2011)

Low-Quality (LQ) Image

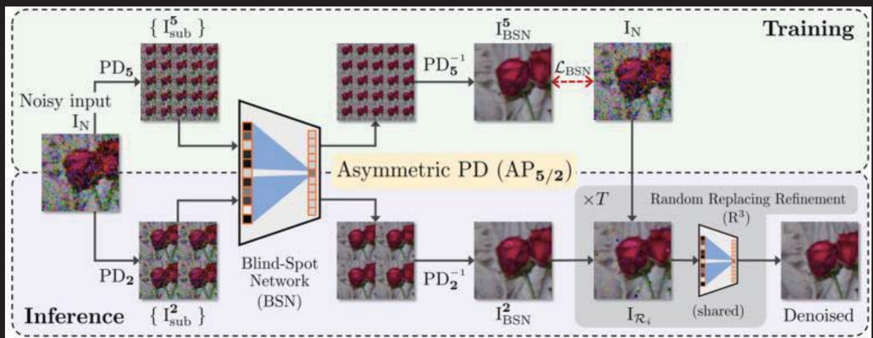
High-Quality (HQ) Image

Previous Works

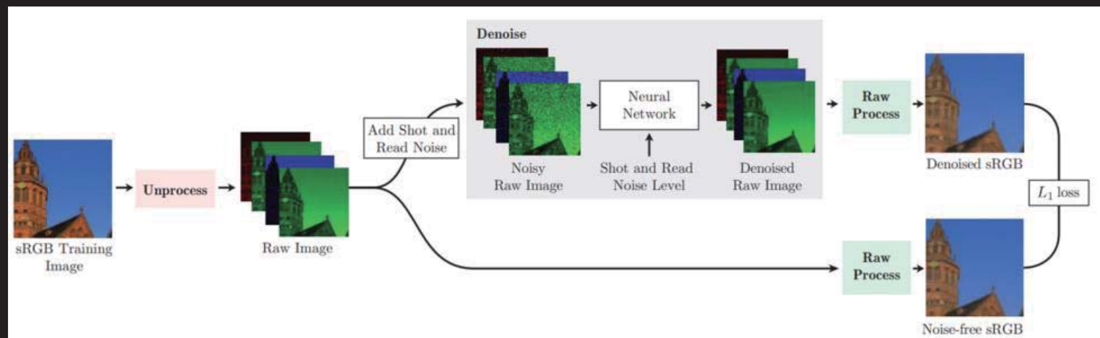
- Typically follows three processes:
  - Prepare training dataset
  - Train DNNs with (self-) supervised learning
  - Generate a single output for a test image input



DPE (CVPR 2018)



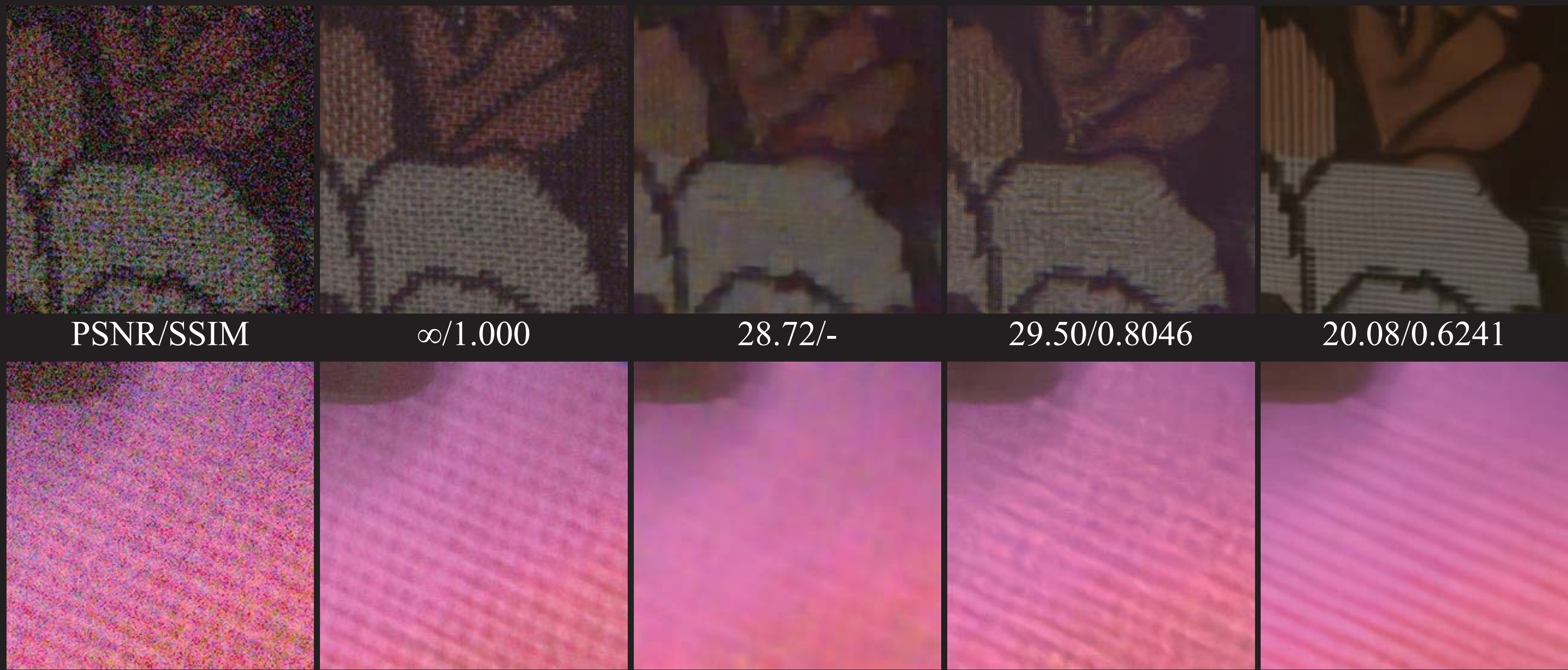
AP-BSN (CVPR 2022)



UPI (CVPR 2019)

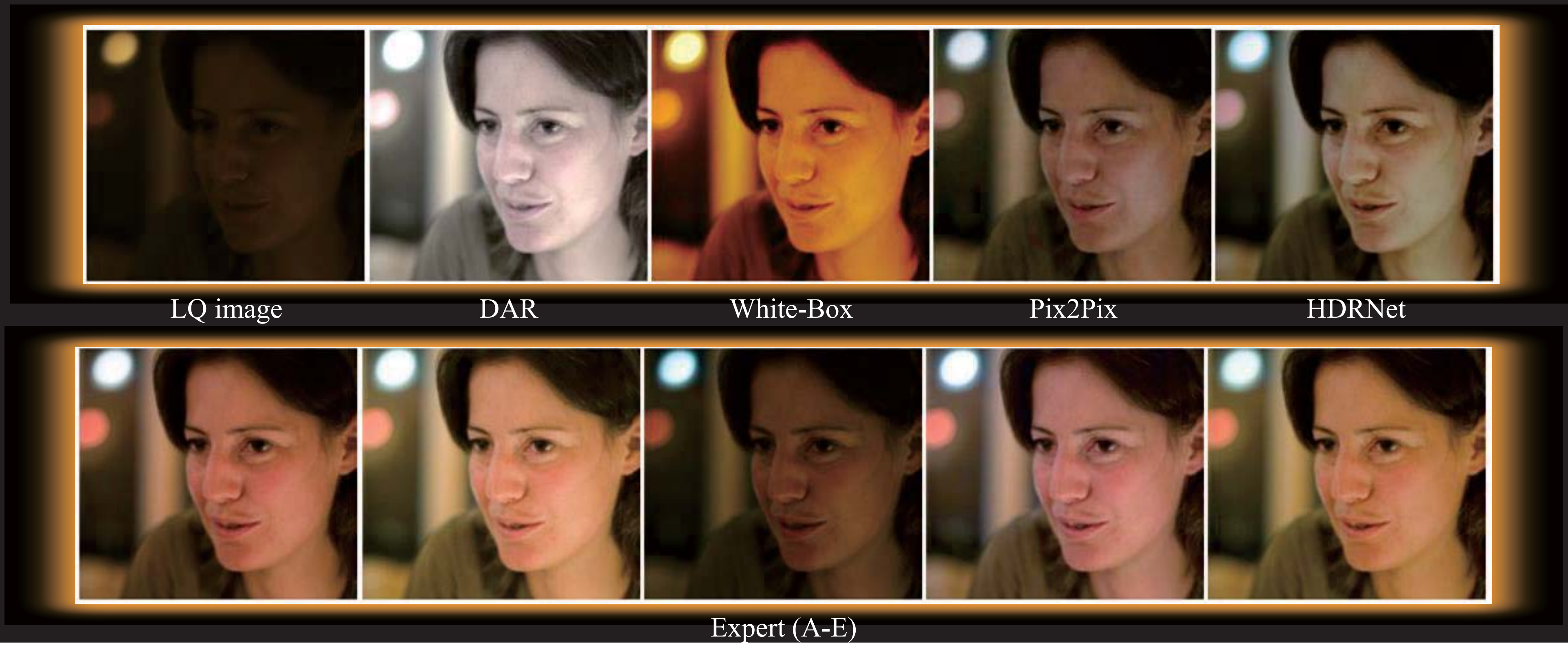
Challenges

- Unknown image degradation → Failure to restore original textures from the degradation of interest



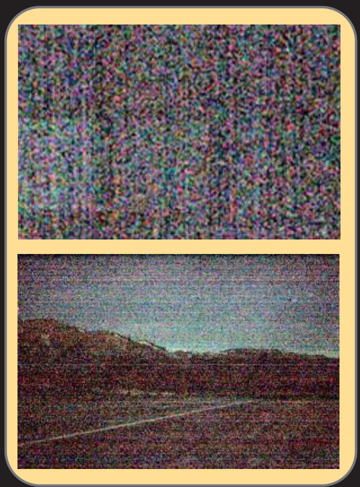
Challenges

- Subjective user preference → Users dissatisfied with the appearance of one image

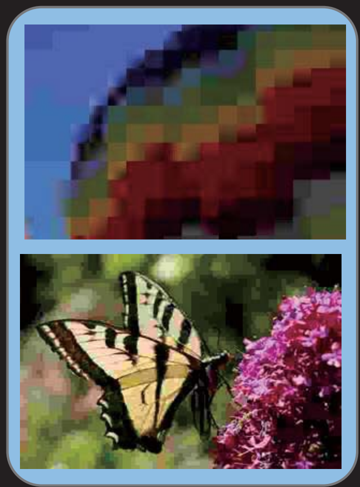


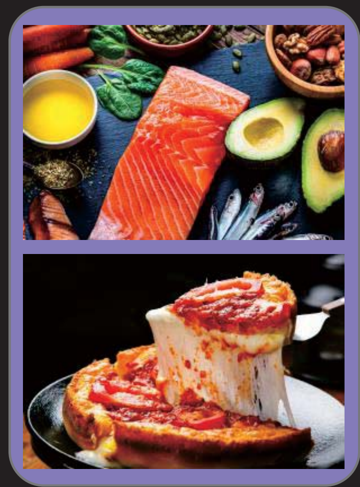
Main Idea: Adaptive Image Signal Processor

- Motivation
  - Digital images contain a wide variety of degradations & Visually pleasing images can have a wide variety of styles
  - Single model with fixed parameters ***cannot generalize*** across all image degradations or all image styles
  - Different models for each degradation or each style are ***computationally infeasible***

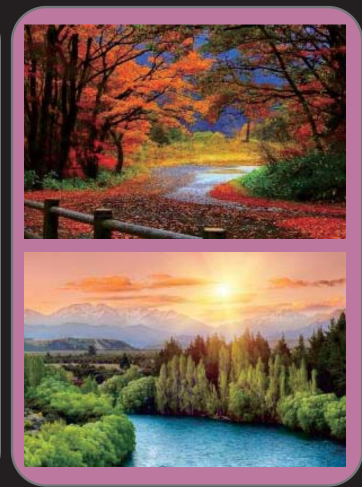

Noise


Blur


Compression artifact


Style for food


Style for low light


Style for nature

Adaptive Image Signal Processor *for* Practical Applications

Adaptive ISP *for* Camera Image Denoising

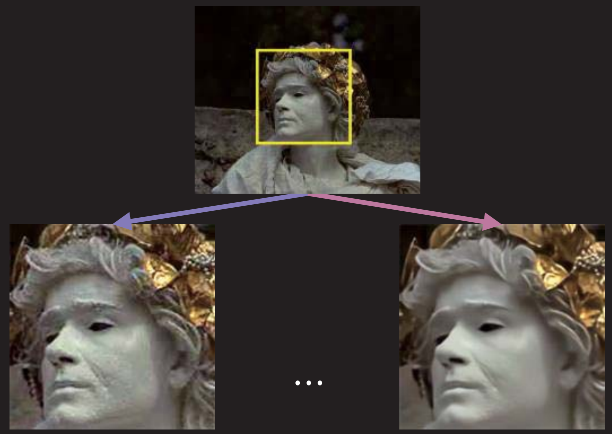
- Camera has a noise model from the image acquisition process of image sensor and RAW2RGB conversion
- The proposed adaptive ISP learns to remove the camera-specific noise through noise estimation & synthesis

Which **noise** is applied?

Adaptive ISP *for* Controllable Image Restoration

- For unknown degradations, a practical application is multiple-output generation for *predetermined* restoration tasks to select a preferred output
- The proposed adaptive ISP generates an output restoration for a task controlled by users

Which **restoration output** is preferred?

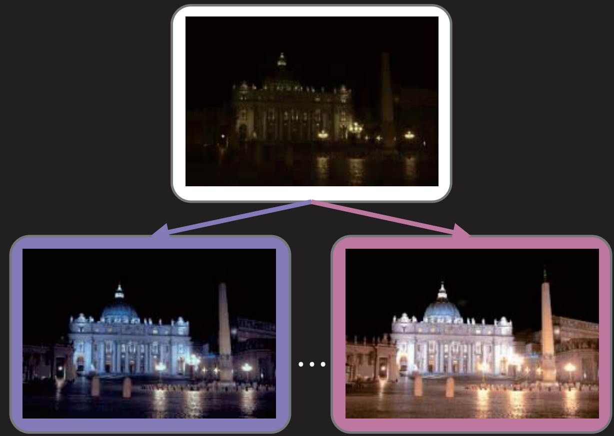


restoration-task-*1* ⋯ restoration-task-*m*

Adaptive ISP *for* Controllable Image Enhancement

- For subjective user preferences, a practical application is multiple-style generation
- The proposed adaptive ISP generates an output image for a style encoded with some user-controllable coefficients

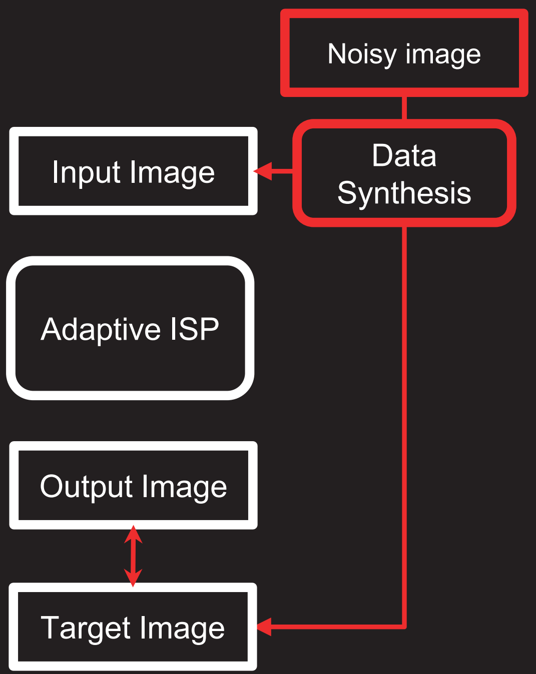
Which **style** is preferred?



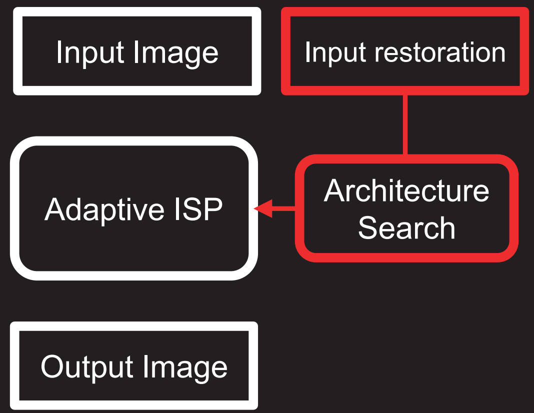
Style-1 ⋯ Style-*m*

Adaptive Image Signal Processor *for* Practical Applications

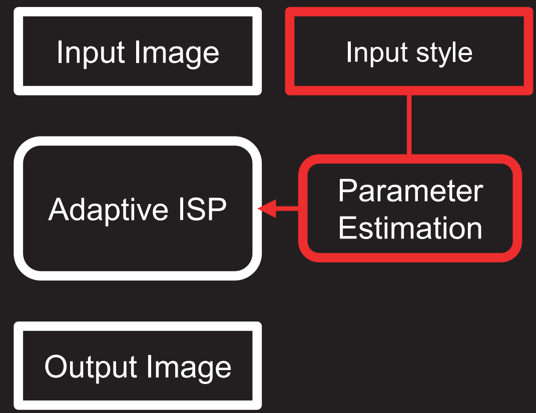
Adaptive Data Synthesis
for Camera Image Denoising



Adaptive Neural Architecture Search
for Controllable Image Restoration



Adaptive ISP Parameter Estimation
for Controllable Image Enhancement



H Kim and KM Lee, NERDS: A General Framework to Train Camera Denoisers from Raw-RGB Noisy Image Pairs, ICLR 2023

H Kim et al., Searching for Controllable Image Restoration Networks, ICCV 2021

H Kim and KM Lee, Learning Controllable ISP for Image Enhancement, TIP 2023

Organization

Chapter 1 | Adaptive Data Synthesis for Camera Image Denoising

Chapter 2 | Adaptive Neural Architecture Search for Controllable Image Restoration

Chapter 3 | Adaptive ISP Parameter Estimation for Controllable Image Enhancement

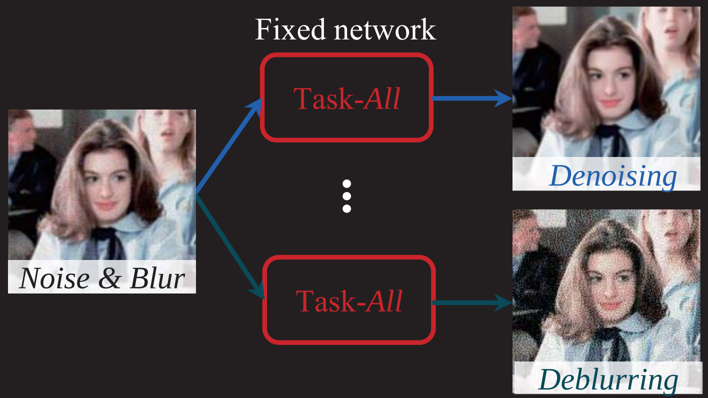
Chapter 2 | Adaptive Neural Architecture Search for Controllable Image Restoration

- *Controllable Image Restoration*
  - Previously not efficient due to a fixed network for all restoration tasks
  - The proposed method greatly improve model efficiency by searching for restoration–task-adaptive architectures

H Kim et al., Searching for Controllable Image Restoration Networks, ICCV 2021

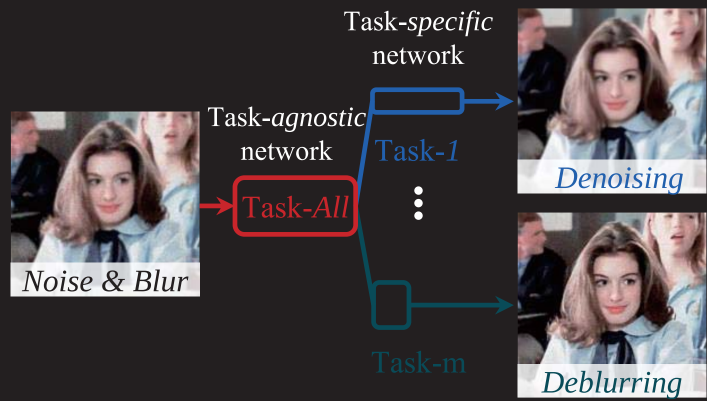
Contributions

- Efficient neural networks for controllable image restoration through *the feature reuse of a task-agnostic network for multiple inferences* and *task-specific architectures*
- A novel supernetwork-based neural architecture search algorithm for controllable image restoration
- Improved output image quality using a new data sampling strategy



$$\mathcal{R}_{\text{total}}(f, \boldsymbol{x}, \boldsymbol{t}) = \sum_{m=1}^{M} \mathcal{R}(f, \boldsymbol{x}, \boldsymbol{t}_m), \tag{1}$$

(a) CResMD (ECCV 2020)



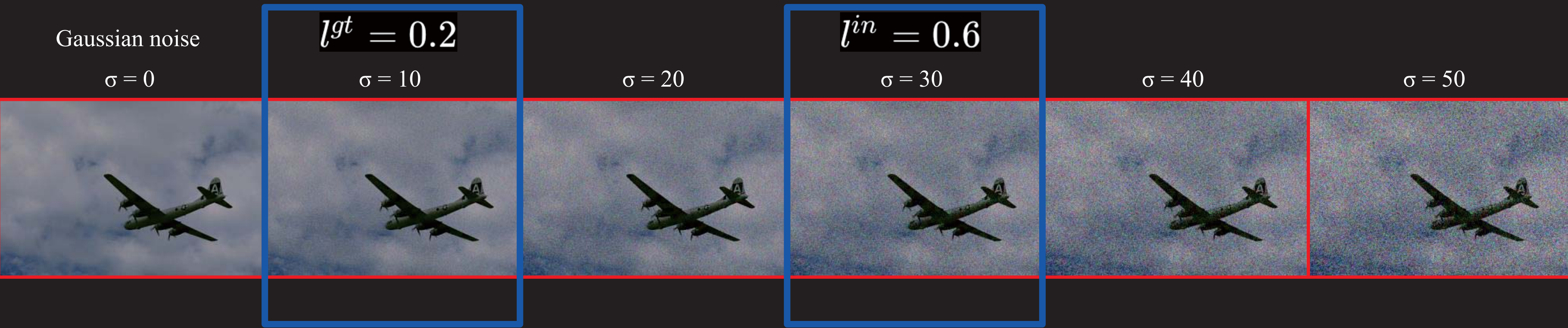
$$\begin{aligned}\mathcal{R}_{\text{total}}(f, \boldsymbol{x}, \boldsymbol{t}) &= \sum_{m=1}^{M} [\mathcal{R}(f^a, \boldsymbol{x}) + \mathcal{R}(f^s, \tilde{\boldsymbol{x}}, \boldsymbol{t}_m)] \\ &\geq \mathcal{R}(f^a, \boldsymbol{x}) + \sum_{m=1}^{M} \mathcal{R}(f^s_m, \tilde{\boldsymbol{x}}, \boldsymbol{t}_m),\end{aligned} \tag{2}$$

(b) TASNet (Ours)

Proposed Method: TASNet (**T**ask-**A**gnostic and Task-**S**pecific **Net**works)

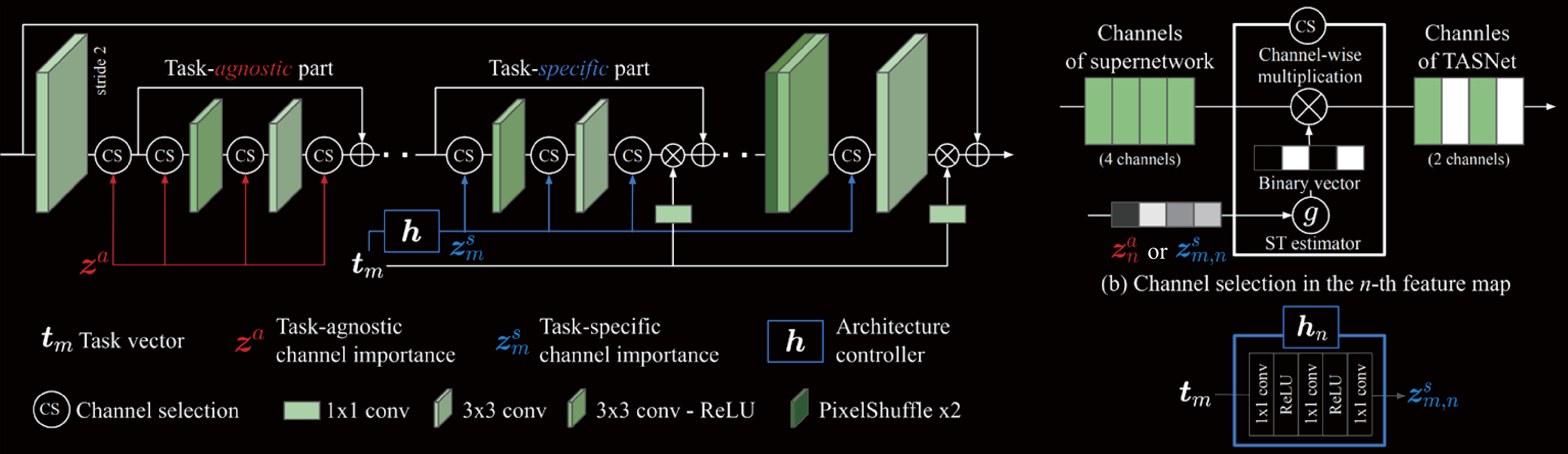
- Definition of restoration tasks with a data sampling strategy (relative GT)
  - Degradation type and level
    - Degradation step : blur ⟶ noise ⟶ compression
    - Degradation level : scaling to [0,1]
  - A restoration task = A task vector
    - Task vector: 3-dimensional vector of degradation differences between input and relative GT for blur, noise, and jpeg
    - Relative GT: Ground truth images with degradations
    - Each dimensional value of task vector becomes $t = l^{in} - l^{gt}$

Degradation Type	level
Gaussian blur	Kernel width ∈ [0,4]
Gaussian noise	σ ∈ [0,50]
Jpeg compression	Quality factor ∈ [10,100], None



Proposed Method: TASNet (**T**ask-**A**gnostic and Task-**S**pecific **Net**works)

- TASNet: *Sharing early layers* across all tasks and *adjusting remaining layers for each task* through *channel selection*



(a) Overview of TASNet architecture

(b) Channel selection in the n-th feature map

(c) Architecture controller in the n-th feature map

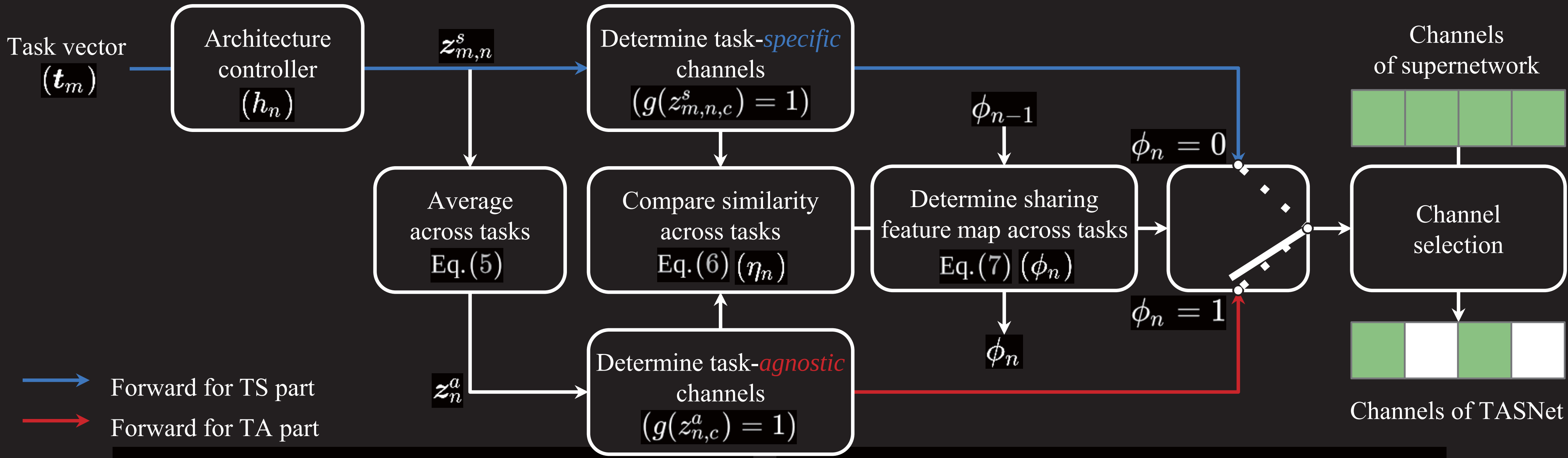
$$g(z) = \begin{cases} \mathbb{I}\,[z > 0] & \text{if forward} \\ \text{sigmoid}(z) & \text{if backward,} \end{cases} \quad (3)$$

$$\boldsymbol{z}^s_{m,n} \equiv h_n(\boldsymbol{t}_m), \quad (4)$$

$$z^a_{n,c} \equiv \frac{1}{M} \cdot \sum_{m=1}^{M} z^s_{m,n,c}, \quad (5)$$

Proposed Method: TASNet (Task-Agnostic and Task-Specific Networks)

- Search process: Determine sharing a feature map across tasks (Eq. (7)), task-*specific* channels, task-*agnostic* channels



$$z^a_{n,c} \equiv \frac{1}{M}\cdot\sum_{m=1}^{M} z^s_{m,n,c}, \tag{5}$$

$$\frac{1}{M}\cdot\sum_{m=1}\sum_{c=1} g(z^s_{m,n,c})\cdot g(z^a_{n,c}) > \gamma\cdot\sum_{c=1} g(z^a_{n,c}), \tag{6}$$

$$\phi_n = \begin{cases} 1 & \text{if } \eta_i = 1, \forall i = 1,2,\ldots,n \\ 0 & \text{otherwise,} \end{cases} \tag{7}$$

$$\min_{\theta,\psi} \mathcal{L}(\theta,\psi) + \lambda_1\cdot\mathcal{R}_1(\psi) + \lambda_2\cdot\mathcal{R}_2(\psi), \tag{8}$$

Proposed Method: TASNet (**T**ask-**A**gnostic and Task-**S**pecific **Net**works)

- Objective functions
  - A standard L1 loss for image restoration with *differentiable resource regularization* terms

$$\min_{\theta,\psi} \mathcal{L}(\theta,\psi) + \lambda_1 \cdot \mathcal{R}_1(\psi) + \lambda_2 \cdot \mathcal{R}_2(\psi), \qquad (8)$$

  - The first resource regularization *penalizes FLOPs* of currently searched architectures by *de-selecting channels*

$$\mathcal{R}_1(\psi) = \mathcal{R}_{\text{FLOPs}}(f^a, \boldsymbol{x}) + \sum_{m=1}^{M} \mathcal{R}_{\text{FLOPs}}(f^s, \tilde{\boldsymbol{x}}, \boldsymbol{t}_m)$$
$$= 2\sum_{n=1}^{N} \mathrm{K}_n^2 H_n W_n \cdot [\phi_n \cdot \sum_{c=1}^{C} g(z^a_{n,c}) \cdot \sum_{c=1}^{C} g(z^a_{n-1,c}) + (1-\phi_n) \cdot \sum_{m=1}^{M} \{\sum_{c=1}^{C} g(z^s_{m,n,c}) \cdot \sum_{c=1}^{C} g(z^s_{m,n-1,c})\}], \quad (9)$$

  - The second regularizer *maximizes the number of shared layers* by *penalizing channel disagreement* across tasks

$$\mathcal{R}_2(\psi) = \sum_{n=1}^{N} \phi_{n-1} \cdot \sum_{c=1}^{C} \sum_{m=1}^{M} \left\| g(z^s_{m,n,c}) - g(z^a_{n,c}) \right\|_1, \quad (10)$$

Experimental Setting

- Dataset
  - Train: DIV2K, 2K 800 images
  - Test: CBSD68,68 RGB HVGA clean images

- Degradation types and levels
  - Degradation step : blur ⟶ noise ⟶ compression
  - Degradation level : scaling to [0,1]
  - Task vector : a 3 dimensional vector

Degradation Type	level
Gaussian blur	Kernel width $\in$ [0,4]
Gaussian noise	$\sigma \in$ [0,50]
Jpeg compression	Quality factor $\in$ [10,100], None

- Measures
  - Reference-based image quality measure: PSNR and SSIM
  - Non-reference-based image quality measure: BRISQUE and NIQE
  - Model efficiency measure: FLOPs, CPU latency, and GPU latency

H Kim et al., Searching for Controllable Image Restoration Networks, ICCV 2021

Experiments

- Computation cost comparisons

Cost metric	Resolution	CResMD	TASNet
FLOPs↓	HD	1,124.3 G	**45.2 G**
	2K	2,698.4 G	**108.4 G**
	4K	10,119.2 G	**406.7 G**
CPU latency (single)↓	HD	22.8 s	**5.5 s**
	2K	55.6 s	**13.5 s**
	4K	209.3 s	**55.5 s**
CPU latency (multi)↓	HD	5.1 s	**1.7 s**
	2K	11.7 s	**4.2 s**
	4K	40.6 s	**13.1 s**
GPU latency↓	HD	144.4 ms	**68.4 ms**
	2K	280.8 ms	**99.2 ms**
	4K	930.0 ms	**250.7 ms**

95.7% reduced FLOPs

2~3 times faster GPU latency

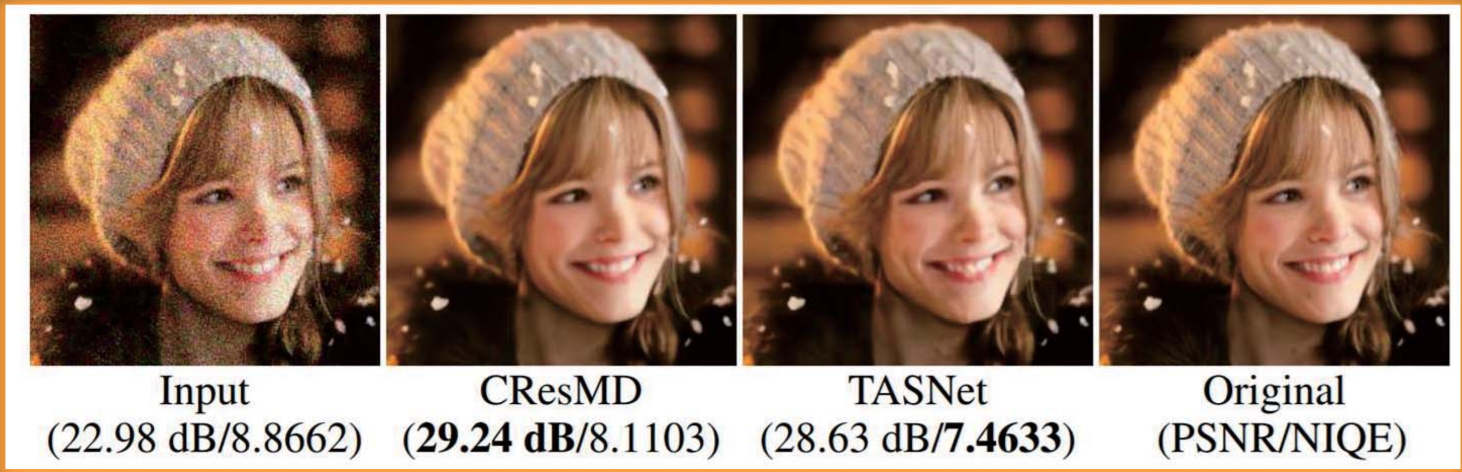
H Kim et al., Searching for Controllable Image Restoration Networks, ICCV 2021

Experiments

- Non-blind setting
  - Quantitative results on CBSD68

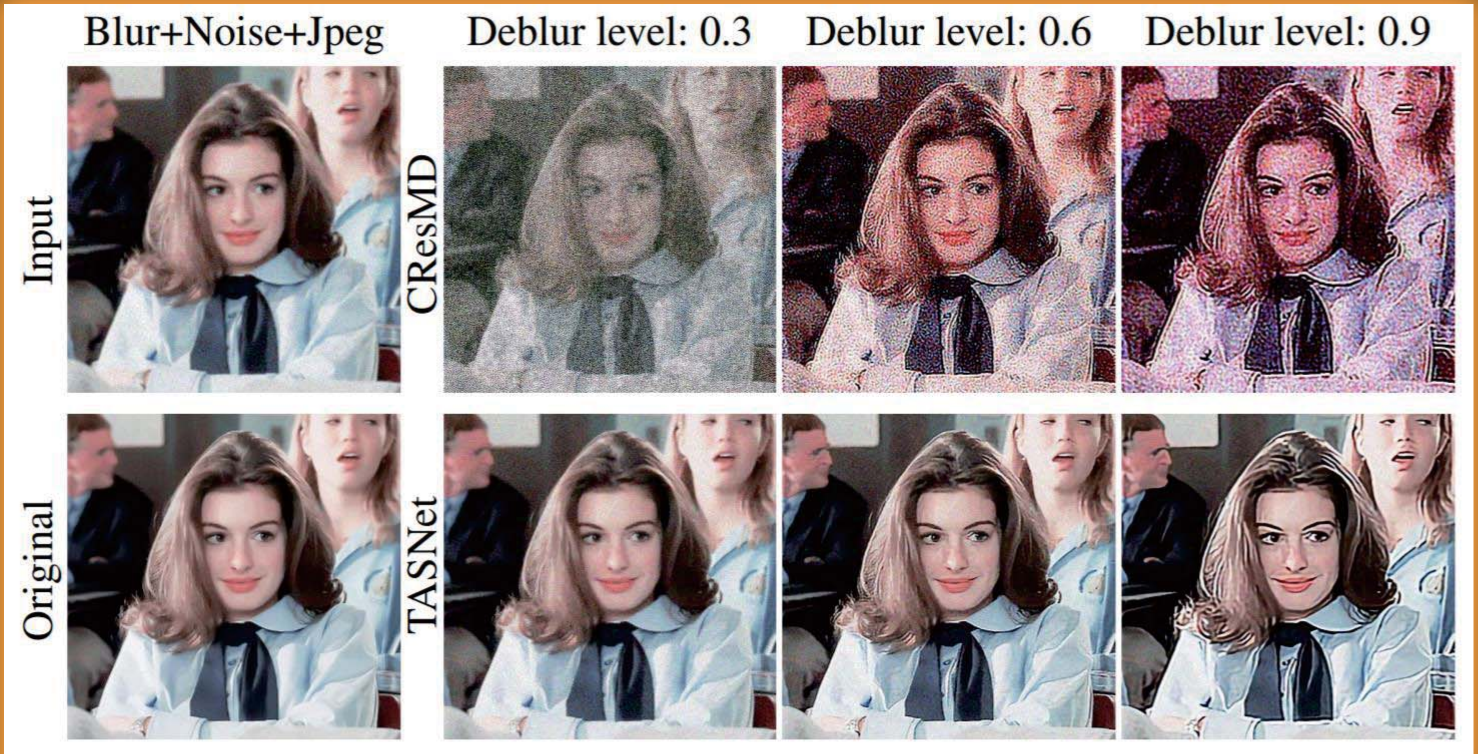
Method	PSNR$_\uparrow$	SSIM$_\uparrow$	NIQE$_\downarrow$	BRISQUE$_\downarrow$	FLOPs$_\downarrow$
CResMD	**25.86 dB**	**0.8194**	6.7165	54.13	189.1 G
TASNet	25.64 dB	0.8137	**6.6301**	**50.60**	**7.5 G**

  - Qualitative results



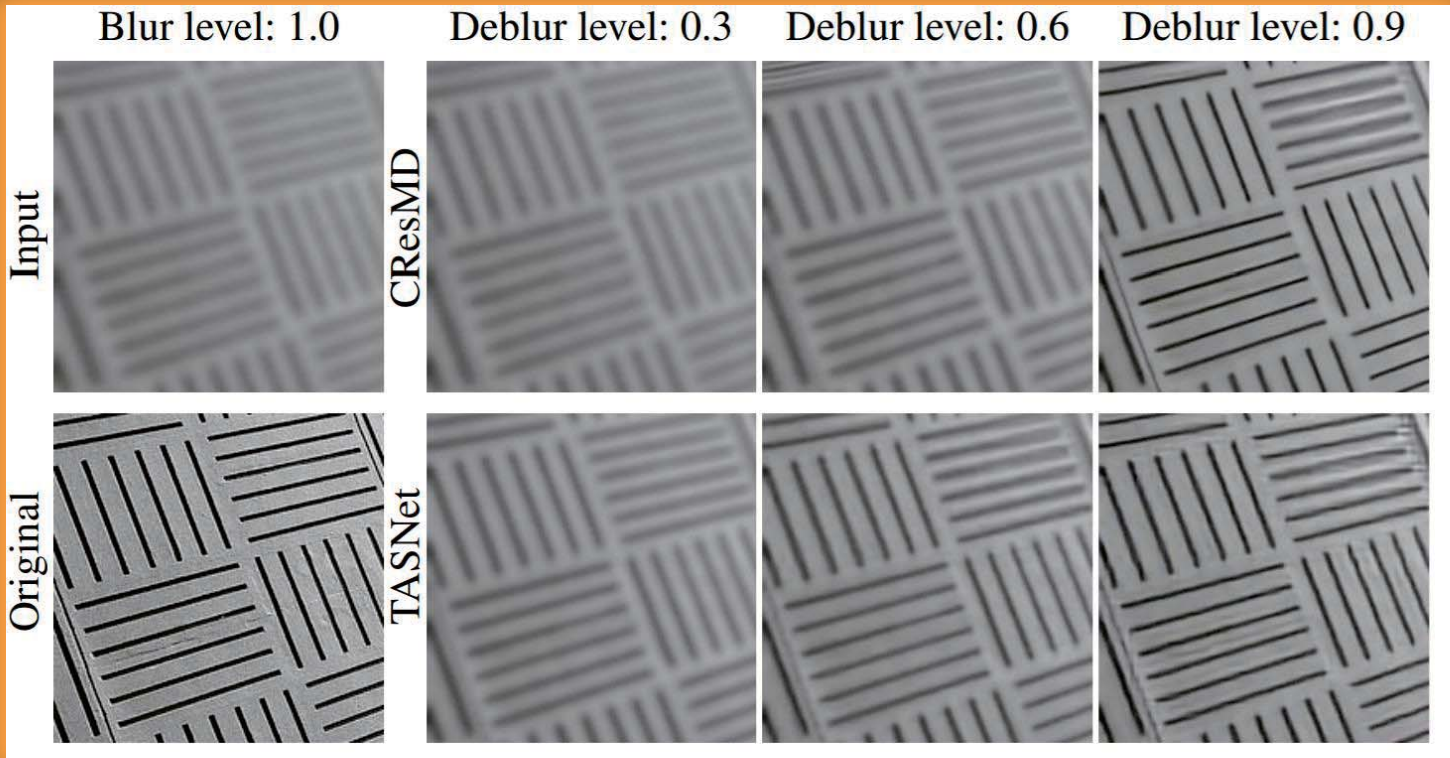
Experiments

- Image quality comparison
  - Deblurring results for blind setting



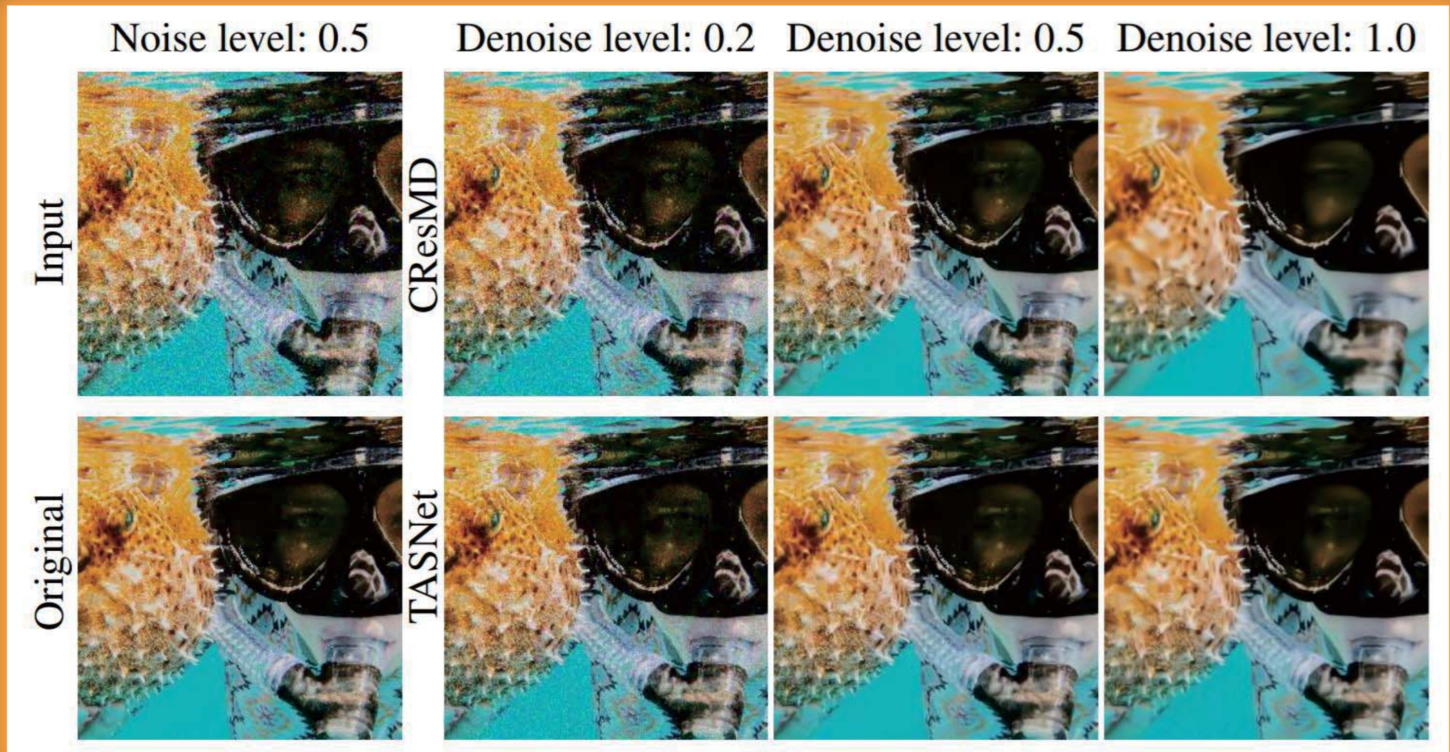
Experiments

- Image quality comparison
  - Deblurring results for blind setting



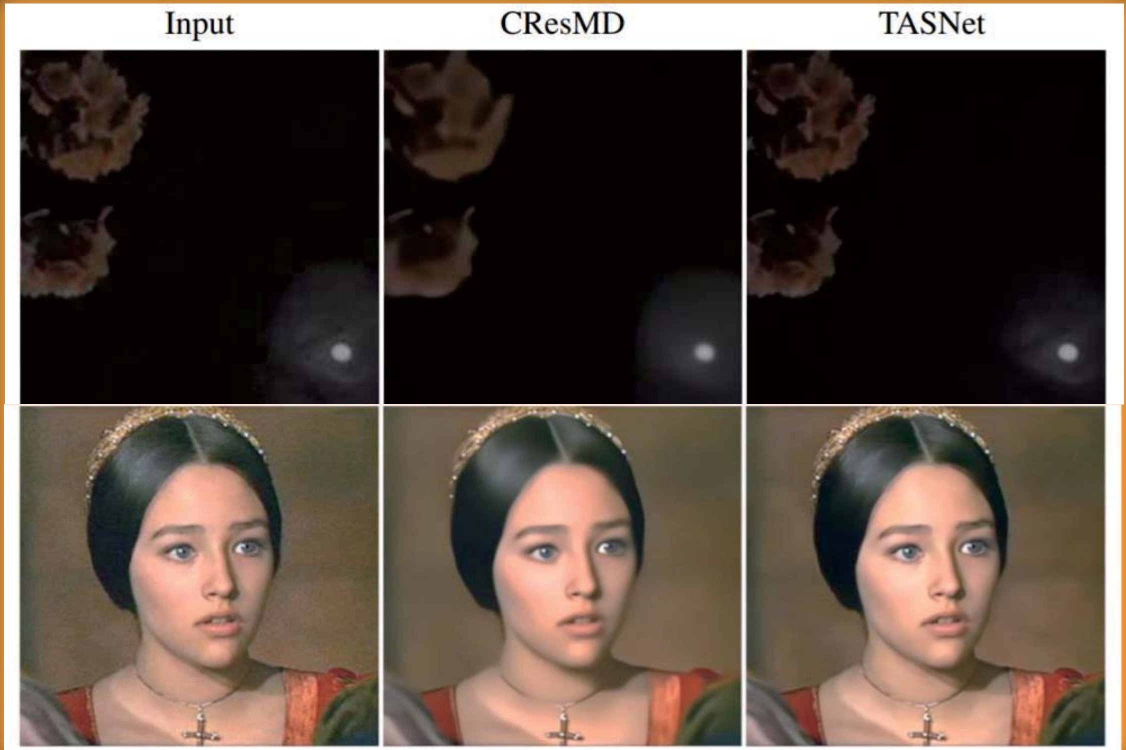
Experiments

- Image quality comparison
  - Denoising results for blind setting



Experiments

- Image quality comparison
  - Restoration results on real images



Experiments

- Image quality comparison
  - NDERDS+D vs. TASNet-Denoising on SIDD validation
    - NERDS+D achieves better PSNR (reference-based) while TASNet achieves better BRISQUE (non-reference-based)

Input | NERDS+D | TASNet-Denoising | GT

PSNR/BRISQUE | **40.63**/69.18 | 28.56/78.21 | 27.48/**44.08** | ∞/54.77

H Kim et al., Searching for Controllable Image Restoration Networks, ICCV 2021

Experiments

- Image quality comparison
  - NDERDS+D & TASNet-Deblurring on SIDD validation
    - TASNet can perform additional deblurring for the results from NERDS+D

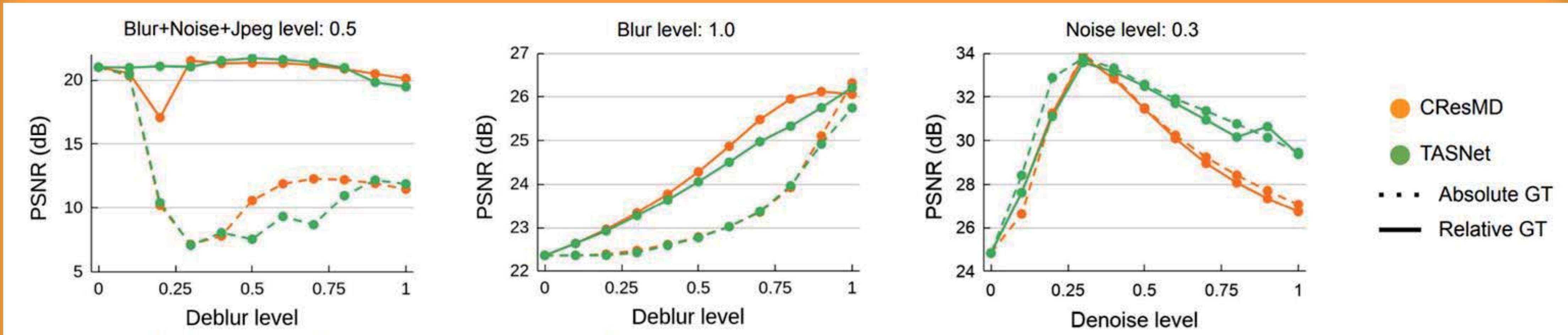
Input	NERDS+D	TASNet-Denoising	NERDS+D & TASNet-Deblurring	GT

PSNR/BRISQUE	**41.74**/66.44	34.17/**36.08**	30.45/43.90	∞/31.77

H Kim et al., Searching for Controllable Image Restoration Networks, ICCV 2021

Experiments

- Analysis
  - Image quality for blind setting

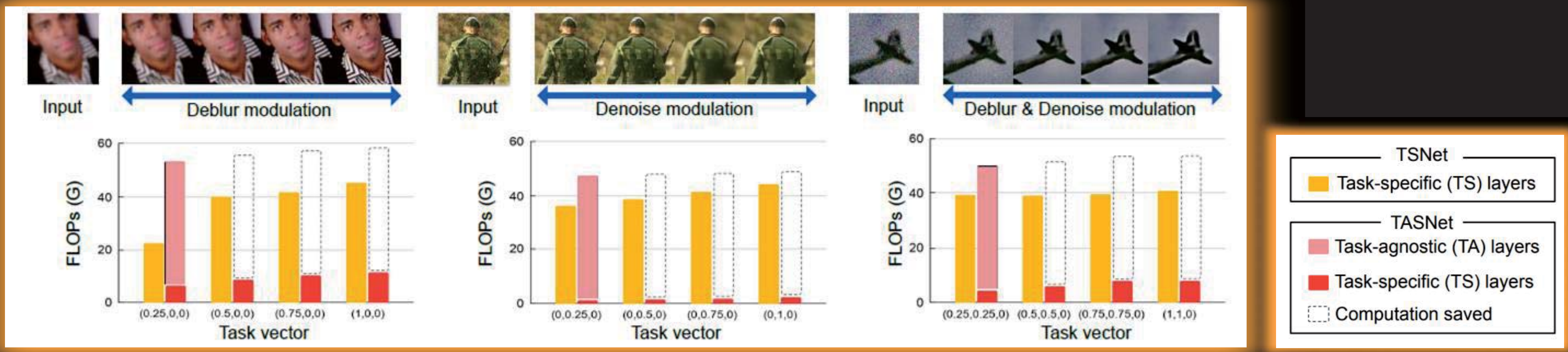


(a) Artifact in generated images (b) Uneven modulation (c) Over-smoothing outputs

H Kim et al., Searching for Controllable Image Restoration Networks, ICCV 2021

Experiments

- Analysis
  - Effectiveness of sharing early layers



H Kim et al., Searching for Controllable Image Restoration Networks, ICCV 2021

Summary

- Adaptive ISP for Controllable Image Restoration through Neural Architecture Search
  - Predetermine general-purpose restoration tasks (deblur, denoise, deartifact)
  - Find TASNet (**T**ask-**A**gnostic and task-**S**pecific **Net**work) through differentiable channel selection from a supernetwork
  - Adapt network architectures and corresponding outputs for the restoration tasks controlled by users
- Flexible Plug-In Module
  - Can perform the general-purpose restoration on real images

H Kim et al., Searching for Controllable Image Restoration Networks, ICCV 2021

Conclusion

- Suggest three independent Adaptive Deep ISPs for Practical Applications
  - Adaptive ISP for camera image denoising through data synthesis
  - Adaptive ISP for controllable image restoration through neural architecture search
  - Adaptive ISP for controllable image enhancement through parameter estimation

- Brought great performance improvement on image quality and model efficiency

- Future works
  - Integration of three proposed methods
  - Advanced practical applications
    - Low-light image denoising & enhancement
    - Dynamic scene deblurring
    - Spatially-varying image enhancement
    - Space telescope ISP