ENCODER-RECURRENT DECODER NETWORK FOR SINGLE IMAGE DEHAZING

An Dang[1], Toan H. Vu[1], Jia-Ching Wang[1,2]

[1]National Central University, Taiwan

[2]Pervasive Artificial Intelligence Research (PAIR) Labs, Taiwan

ICASSP2020
Barcelona

Outline

- Introduction
- The proposed method
- Experiments
- Conclusions

Introduction

- Haze affect visible quality
- Single image dehazing
- Applications in visual systems

(a) Haze

(b) Ours

The atmospheric scattering model (1)

- A degrade hazy image is formulated from its corresponding clear version as

$$\boldsymbol{I}(\boldsymbol{x}) = \boldsymbol{J}(\boldsymbol{x})\boldsymbol{t}(\boldsymbol{x}) + \boldsymbol{A}(\mathbf{1} \; - \boldsymbol{t}(\boldsymbol{x})) \quad (1)$$

Where $I(x)$: the observed hazy image; $J(x)$: the clear image;

x: pixel location. $t(x)$: the transmission map; A: atmospheric light.

- If the atmosphere is homogeneous, we have $t(x) = \; e^{-\beta d(x)}$

Where $d(x)$: the scene depth, β: the scattering coefficient.

The atmospheric scattering model (2)

- (1) is an ill-posed problem as only $I(x)$ is observed.

→ Previous works: recover J(x) by estimating t(x) and A.

- Prior-based approaches like dark channel prior [7], color attenuation [4] → work under restricted assumptions.
- Some DL models [8, 9] directly learn unknown components in the physical model → Performance is limited due to the assumption of an identical atmosphere (in fact, $A = A(x),\ \beta = \beta(x)$)

The proposed method

- Single image dehazing → image-to-image translation
- Don't rely on the atmospheric scattering model
- Encoder-Recurrent Decoder Network (ERDN)
  - Encoder: introduce residual efficient spatial pyramid (rESP) module as a main component to extract multi-level features.
  - Decoder: present the use of ConvRNN to aggregate the encoded features to recover the clear image.

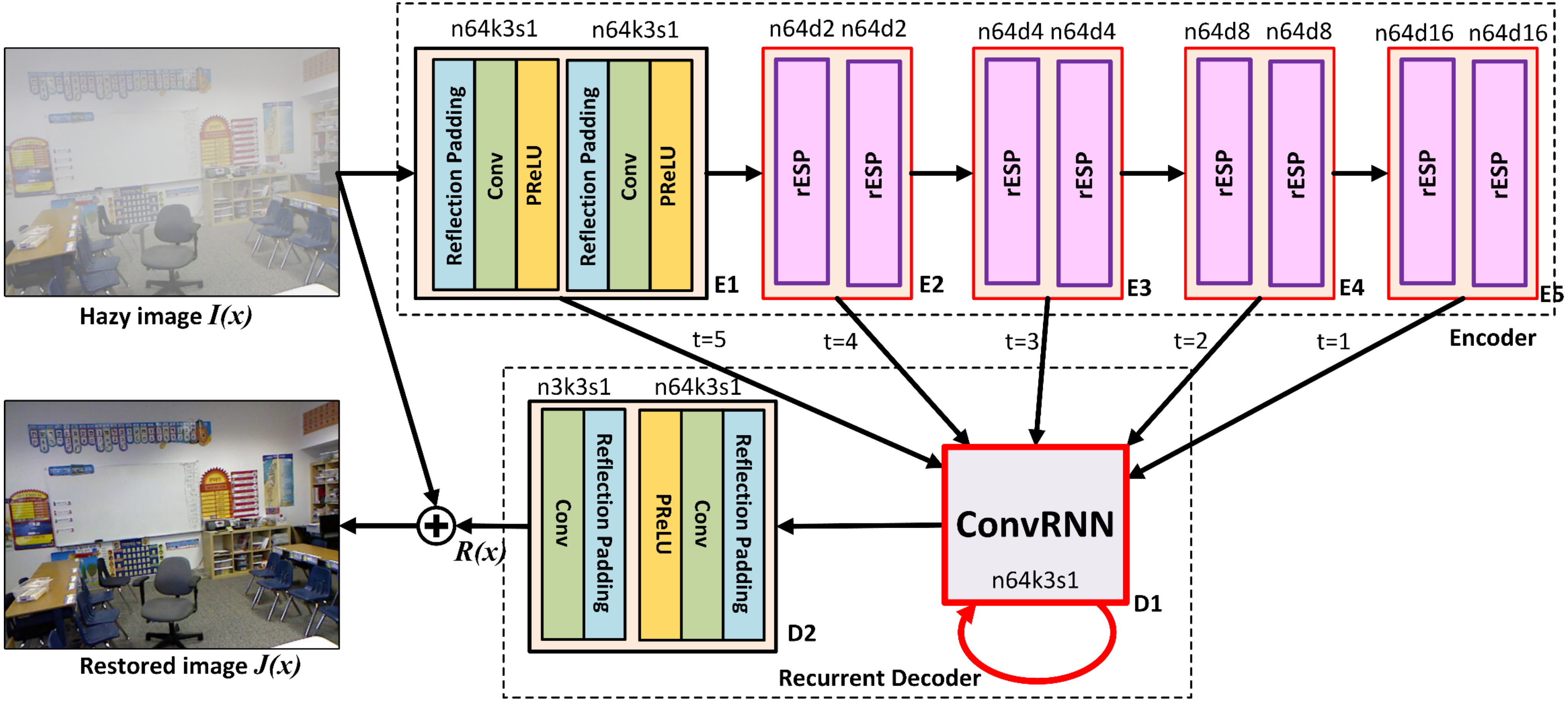
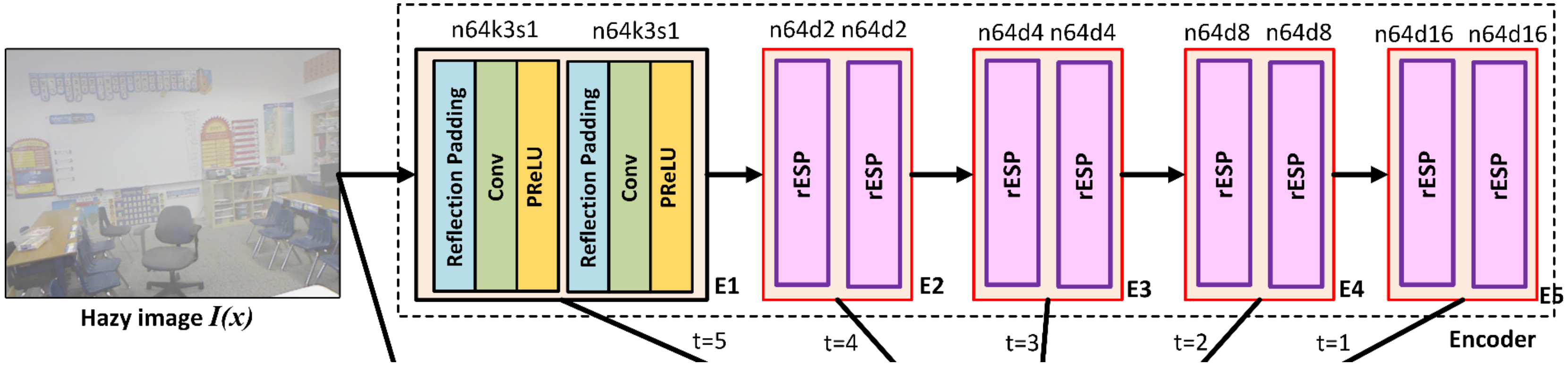


Fig. 1: The proposed encoder-recurrent decoder network (ERDN). ERDN includes two parts- an encoder as the upper branch, and a recurrent decoder as the lower branch. n, k, s, and d denote number of output channels, kernel size, stride, and dilation rate, respectively. In details, we apply $n = 64, k = 3, s = 1$ for all layers, excepts at output we have $n = 3$.

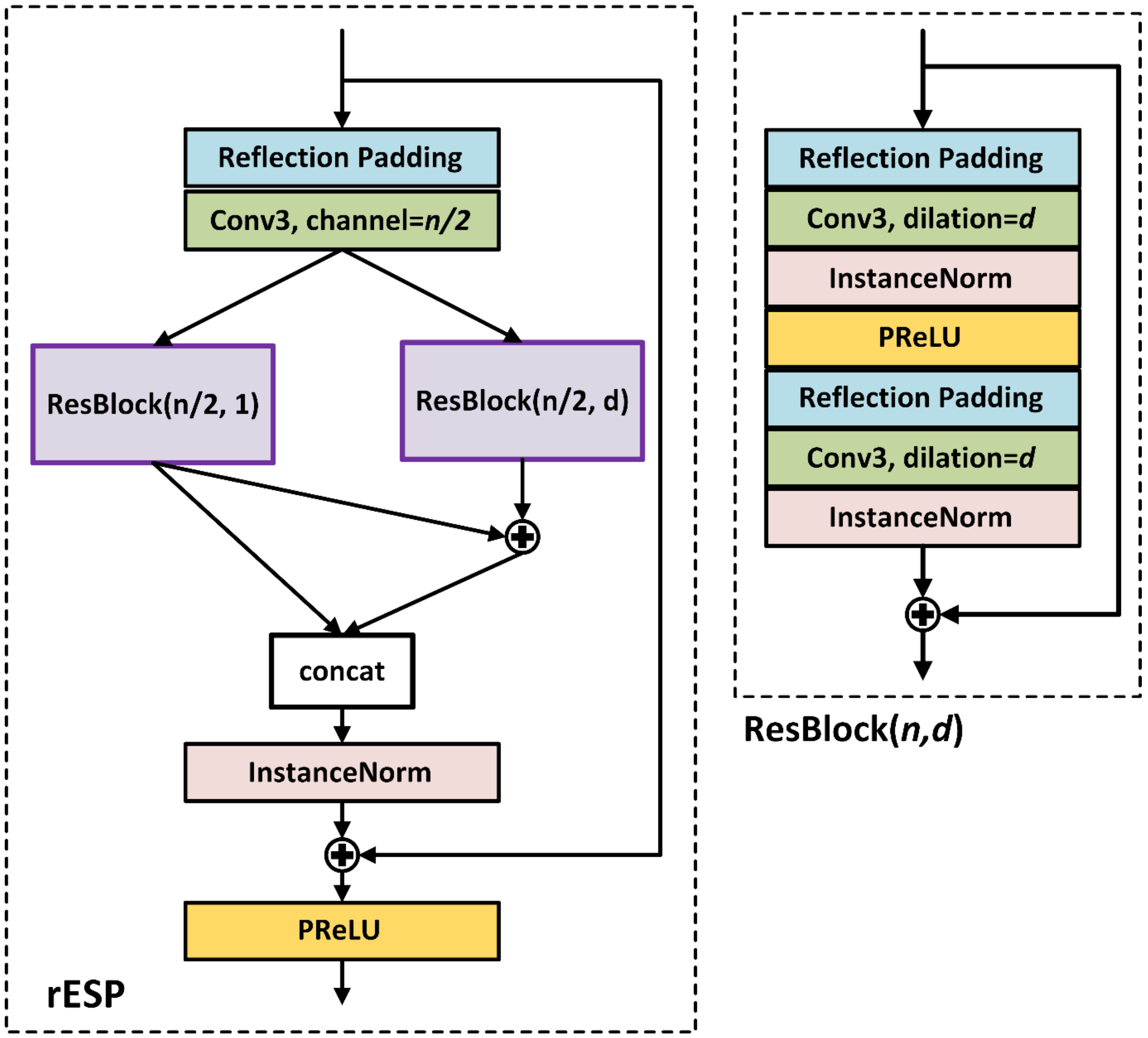
The Encoder

- Combine local features and global features
  - Local features present local information such as textures, shape, and color.
  - Global features provide contextual information
- Residual efficient spatial pyramid (rESP) module
- Consist of 1 convolutional block + 4 rESP blocks



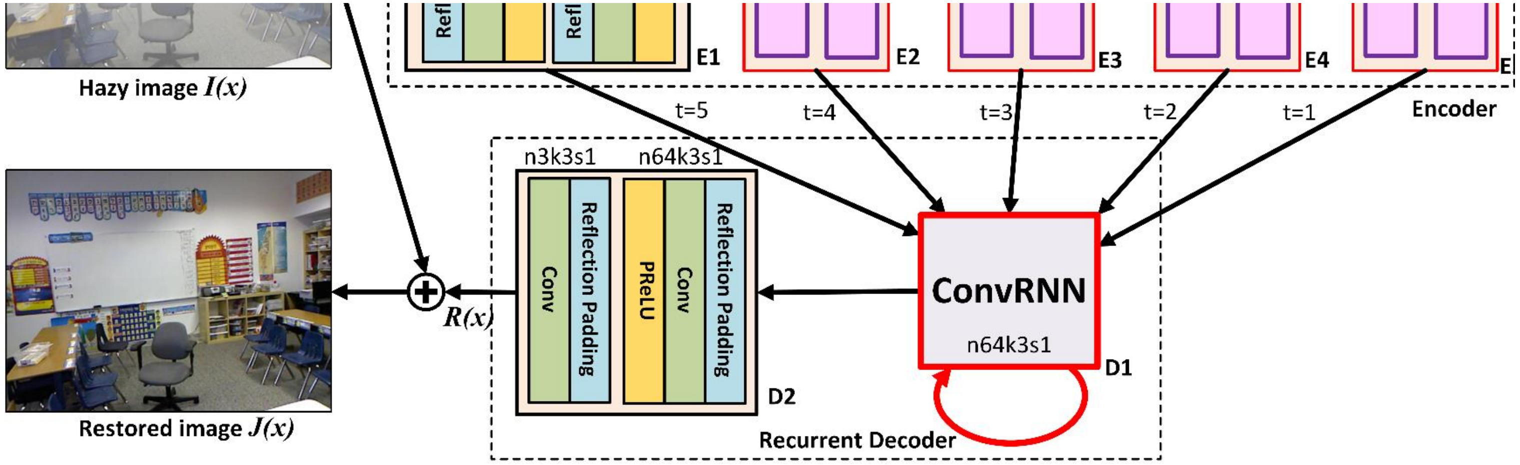
rESP

- Integrates dilated resblock [15, 19] into the ESP module [16]
- Dilated resblock helps to enlarge receptive fields quickly in a few layers.
- The mechanism of feature fusion in the ESP module smooths the effect of large dilation rates →reduce the gridding artifacts [15].



The Recurrent Decoder

- In previous works: a decoder in the U-Net style (# of blocks in the decoder is similar to that of the encoder) →increases model size and computation
- The use of ConvRNN to sequentially aggregate the encoded features from high levels to low levels.
- Specifically, a convolutional control gate-based recurrent neural network (ConvCGRNN) [17] is developed → effective and efficient.



ConvCGRNN

The temporal state

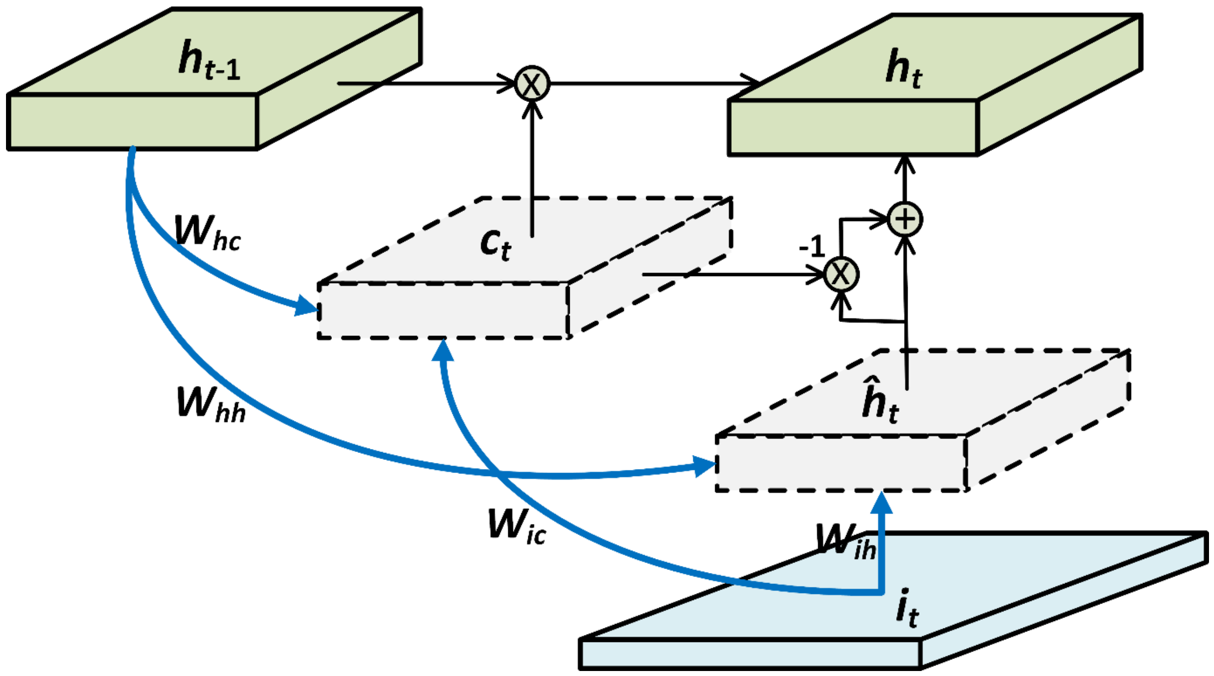
$$\hat{h}_t = f(W_{ih} * i_t + W_{hh} * h_{t-1})$$

The control gate

$$c_t = \sigma(W_{ic} * i_t + W_{hc} * h_{t-1})$$

The new hidden state

$$h_t = c_t \otimes h_{t-1} + (1 - c_t) \otimes \hat{h}_t$$



Experiments

- RESIDE-Standard dataset
- Two sets: an indoor training set (ITS), and a synthetic objective testing set (SOTS)
- The ITS consists of 13990 hazy images generated from 1399 clear images
  - Split ITS to two parts train/validation: 80/20 for training
- The SOTS comprises two parts: indoor and outdoor, each has 500 images
  - Indoor set for in-domain evaluation
  - Outdoor set for cross-domain evaluation

Training details

- MSE loss function $L = \frac{1}{N}\sum_{k=1}^{N} ||J_k - O_k||^2$

- Data augmentation:

  - Randomly rotate and crop images at size of (256 × 256)

  - Creating new synthetic hazy images (*RandomFog* function of *Albumentations* [22])

- During training, scan for a set of difficult examples → more training time.

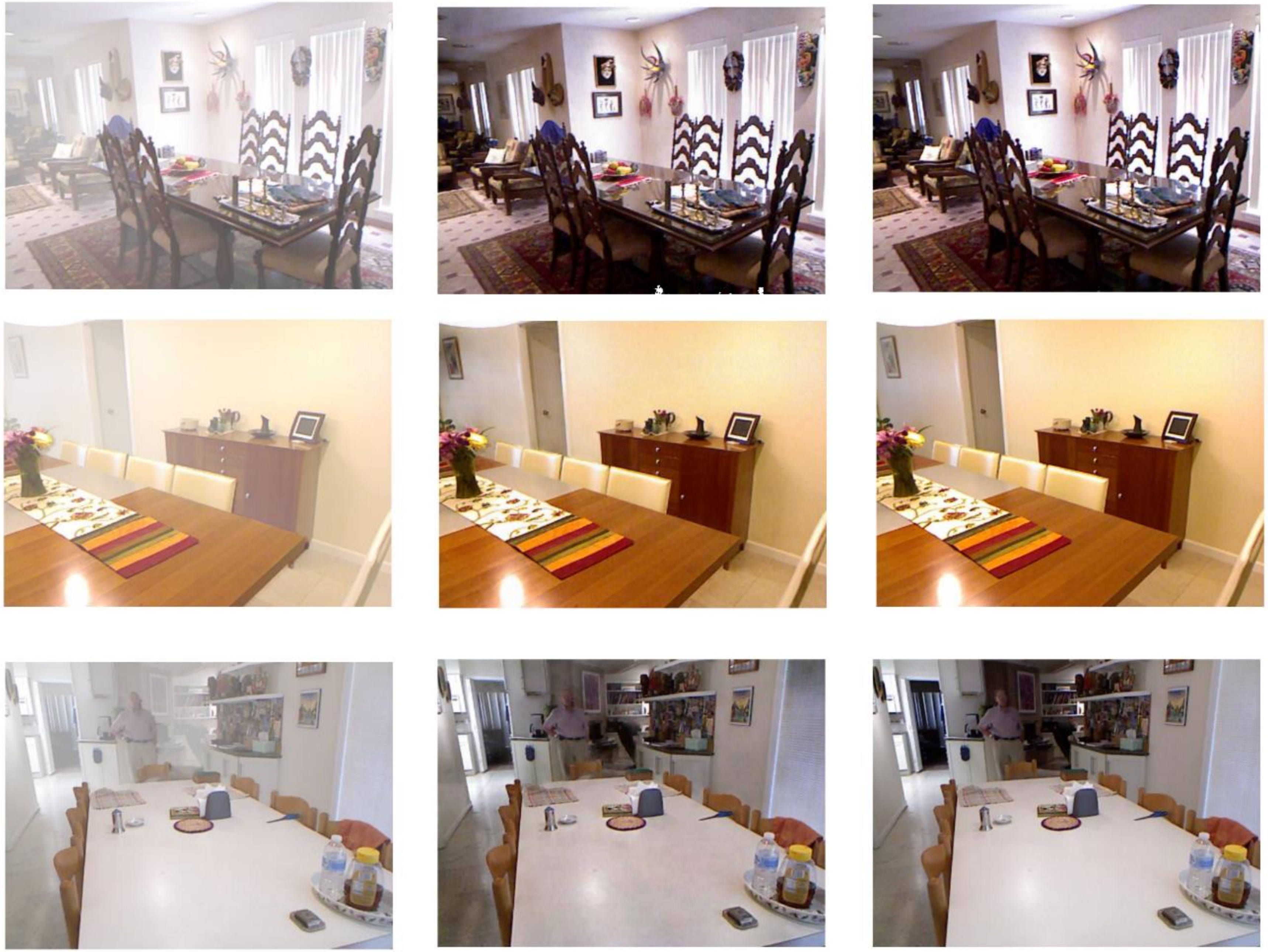
Evaluation results

Table 1: Performance on RESIDE-Standard SOTS Indoor dataset.

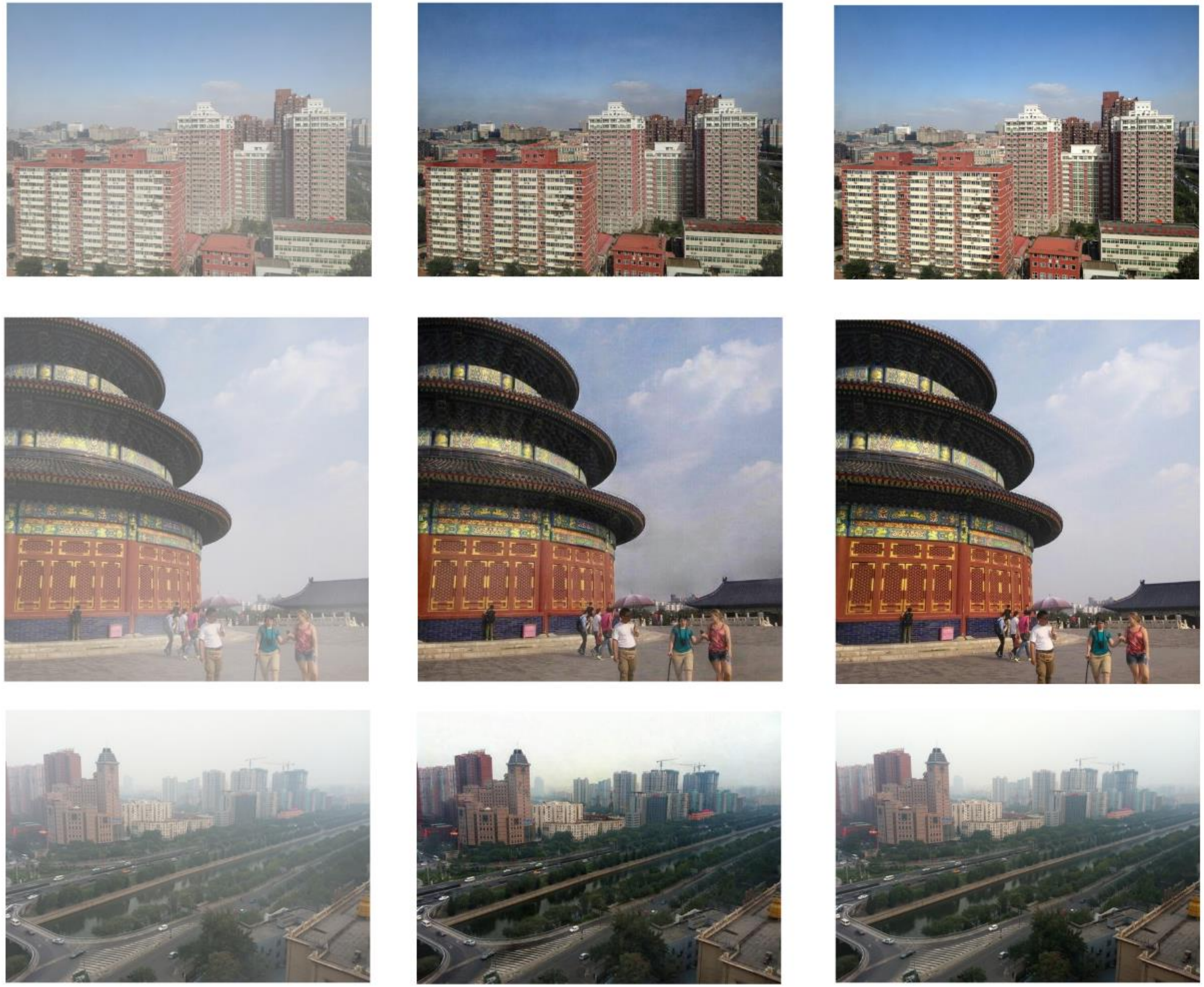
Metrics	DCP [7]	CAP [4]	NLD [5]	DehazeNet [8]
PSNR SSIM	16.62 0.8179	19.05 0.8364	17.29 0.7489	21.14 0.8472
AOD-Net [14]	GMAN [10]	E.Pix2pix [12]	U-net [11]	Ours
19.06 0.8504	<u>27.94</u> 0.897	25.06 0.9232	27.79 **0.9556**	**28.14** <u>0.9522</u>

Table 2: Cross-domain evaluation on RESIDE-Standard SOTS Outdoor dataset.

Metrics	DCP [7]	DehazeNet [8]	AOD-NET [14]	DCPDN [9]
PSNR SSIM	19.13 0.8148	22.46 0.8514	20.29 <u>0.8765</u>	19.93 0.8449
	GFN [13]	E.Pix2pix [12]	Ours	
PSNR SSIM	21.55 0.8444	<u>22.57</u> 0.8630	**24.15** **0.8975**	



(a) Hazy images (b) Our results (c) Ground-truth images



(a) Hazy images (b) Our results (c) Ground-truth images

Ablation study

(1) Different number of blocks in the encoder

(2) Changing rESP block by ResBlock and ESP block

(3) Changing ConvCGRNN by ConvVRNN, ConvGRU and ConvLSTM

Table 3: Ablation study on RESIDE-Standard SOTS Indoor dataset.

Metrics	E1-3	E1-4	**Full** (E1-5, rESP block, ConvCGRNN)
PSNR	25.55	26.49	
SSIM	0.9162	0.9425	
	ResBlock	ESP block	
PSNR	**28.36**	26.61	<u>28.14</u>
SSIM	<u>0.9472</u>	0.9361	**0.9522**
	ConvVRNN	ConvGRU	ConvLSTM
PSNR	26.94	27.98	27.35
SSIM	0.9347	0.9424	0.9416

Conclusions

- Encoder-recurrent decoder network for single image dehazing problem
- Newly introduce the use of two components for the model construction
  - Residual efficient spatial pyramid (rESP)
  - ConvCGRNN
- The ERDN demonstrates its effectiveness and efficiency on the problem.

Thank you very much