Code and Paper

Chengyu's Homepage

NEURAL INFORMATION PROCESSING SYSTEMS

Real-world Image Dehazing with Coherence-based Label Generator and Cooperative Unfolding Network

Chengyu Fang, Chunming He†, Fengyang Xiao, Yulun Zhang†, Longxiang Tang, Yuelin Zhang, Kai Li, Xiu Li†

清華大學 Tsinghua University

Duke UNIVERSITY

上海交通大學 SHANGHAI JIAO TONG UNIVERSITY

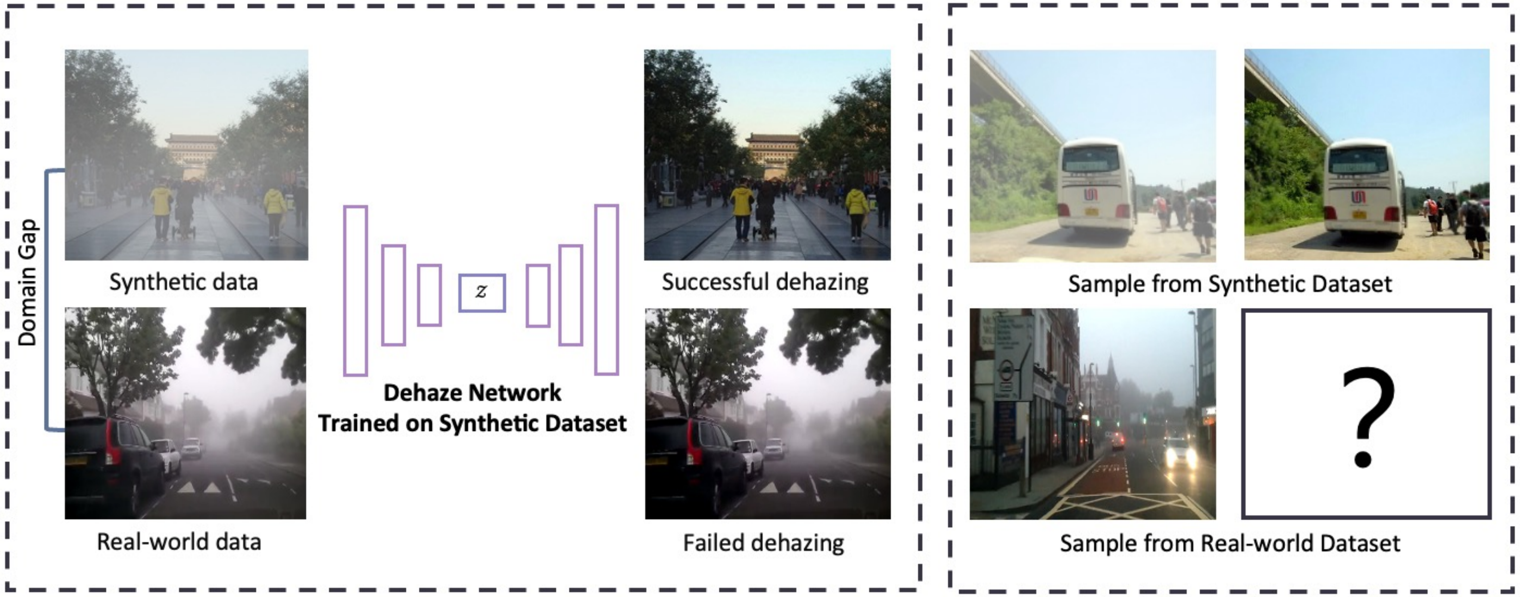
中山大學 SUN YAT-SEN UNIVERSITY

香港中文大學 The Chinese University of Hong Kong

Meta

NEURAL INFORMATION PROCESSING SYSTEMS

The real-world image dehazing task remains challenging due to the **complexities in accurately modeling real haze distributions** and **the scarcity of paired real-world data**.



"The domain gap between synthetic data and real-world data is like an ocean, separating us from each other."

NEURAL INFORMATION PROCESSING SYSTEMS

To address these challenges, our contributions are summarized as follows:

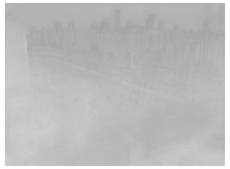
- We propose a **dehazing method**, the COopeRative Unfolding Network (CORUN)
- We propose a semi-supervised domain-adaptation **framework**, Coherence-based Pseudo-label Generator (Colabator)
- We evaluate our **CORUN** with the **Colabator** framework on real-world dehazing tasks. Abundant experiments demonstrate that our method achieves state-of-the-art performance.

“Don’t worry, with Colabator’s help, even the widest domain gap can feel as close as a dear friend from afar, bridging the distance with ease.”

The atmospheric scattering model (ASM): $P(x) = J(x)t(x) + A(1 - t(x))$

Hazy image | Clear image | Atmospheric light | Transmission map

{R,G,B}



Some of previous ASM-based Methods: Estimate A and $t(x)$ then calculate: $$J(x) = \frac{P(x) - A(1 - t(x))}{t(x)}$$

{R,G,B}

- Estimating atmospheric light and the transmission map separately ignores the correlated features between them.
- It ignores the diversity of degradation in real-world scenes beyond hazy.

Cooperative Unfolding Network

- **We implicitly estimate A to focus on the detailed characterization of the scene and the relationship between volumetric haze and scene:**

$$P(x) = J(x)t(x) + A(1 - t(x)) \Longrightarrow \mathbf{P} = \mathbf{J} \cdot \mathbf{T} + \mathbf{I} - \mathbf{T}$$

Where $\mathbf{P}$:  {R,G,B} $\mathbf{J}$:  {R,G,B} $\mathbf{I}$: all-one matrix $\mathbf{T}$: Transmission Map

: Scene

- **Based on the simplify formulation, we can define our cooperative dehazing energy function like:**

$$L(\mathbf{J}, \mathbf{T}) = \frac{1}{2}\|\mathbf{P} - \mathbf{J} \cdot \mathbf{T} + \mathbf{T} - \mathbf{I}\|_2^2 + \psi(\mathbf{J}) + \phi(\mathbf{T})$$

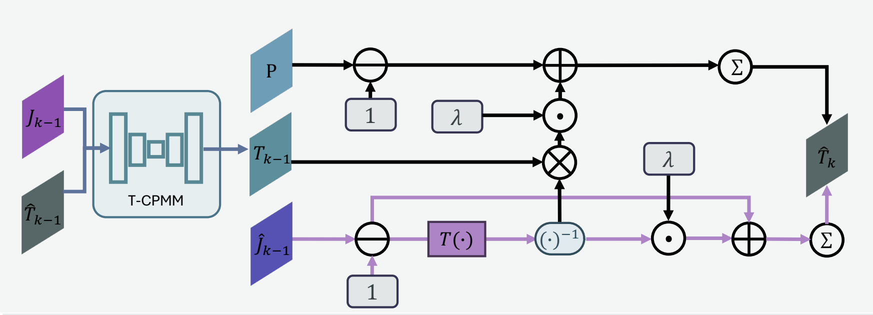
Where $\psi(\mathbf{J})$ and $\phi(\mathbf{T})$ are regularization terms on $\mathbf{J}$ and $\mathbf{T}$.

- **We introduce two auxiliary variables $\hat{\mathbf{T}}$ and $\hat{\mathbf{J}}$ to approximate $\mathbf{T}$ and $\mathbf{J}$, respectively. This leads to the following minimization problem:**

$$\{\hat{\mathbf{J}}, \hat{\mathbf{T}}\} = \arg\min_{\mathbf{J}, \mathbf{T}} L(\mathbf{J}, \mathbf{T})$$

Now, let's optimize the transmission and scene based on PGD algorithm and our cooperative deep unfolding network. For Transmission Optimization:

NEURAL INFORMATION PROCESSING SYSTEMS



(ii) Details of TGDM at the k^{th} stage.

- Give the estimated coarse transmission map T and dehazed image $\hat{\mathrm{J}}_{k-1}$ at iteration $k-1$, the variable T can be updated as:

$$\mathbf{T}_k = \arg\min_{\mathbf{T}} \frac{1}{2} \left\| \mathbf{P} - \hat{\mathbf{J}}_{k-1} \cdot \mathbf{T} + \mathbf{T} - \mathbf{I} \right\|_2^2 + \phi(\mathbf{T})$$

- We construct the proximal mapping between $\hat{\mathrm{T}}$ and T by a encoder-decoder like neural network which we named T-CPMM and denoted as $prox_{\phi}$:

$$\mathbf{T}_k = \mathrm{prox}_{\phi}(\mathbf{J}_{k-1}, \hat{\mathbf{T}}_k)$$

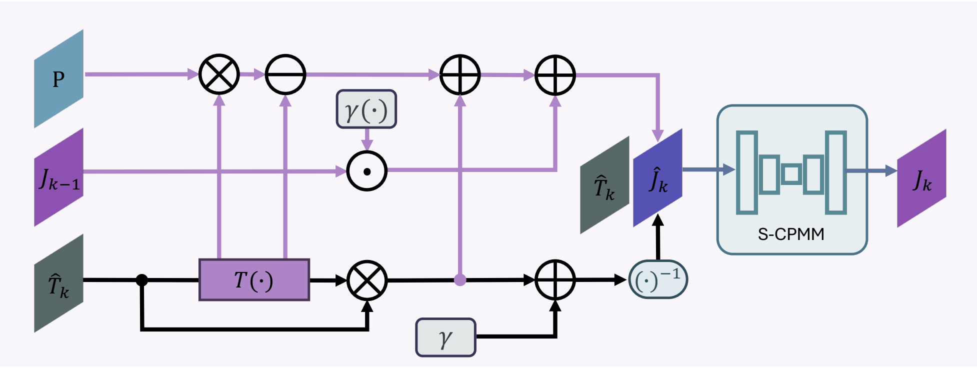
- The auxiliary variables $\hat{\mathrm{T}}$, which we calculate by our proposed TGDM can be formulated as:

$$\hat{\mathbf{T}}_k = \sum_{c \in \{R,G,B\}} (\mathbf{I} - \hat{\mathbf{J}}_{k-1}^c + \lambda_k (\mathbf{I} - \hat{\mathbf{J}}_{k-1}^c)^{-\top})^{-1} \cdot (\mathbf{I} - \mathbf{P}^c + \frac{\lambda_k \mathbf{T}_{k-1}}{(\mathbf{I} - \hat{\mathbf{J}}_{k-1}^c)^{\top}})$$

The variable λ_{k} is a learnable parameter, we learn this parameter at each stage during the end-to-end learning process, allowing the network to adaptively control the updates in iteration.

NEURAL INFORMATION PROCESSING SYSTEMS

For Scene Optimization:



(iii) Details of SGDM at the k^{th} stage.

- Give $\hat{\mathrm{T}}_k$ and J, the variable J can be updated as:

$$\mathbf{J}_k = \arg\min_{\mathbf{J}} \frac{1}{2} \|\mathbf{P} - \mathbf{J} \cdot \hat{\mathbf{T}}_k + \hat{\mathbf{T}}_k - \mathbf{I}\|_2^2 + \psi(\mathbf{J})$$

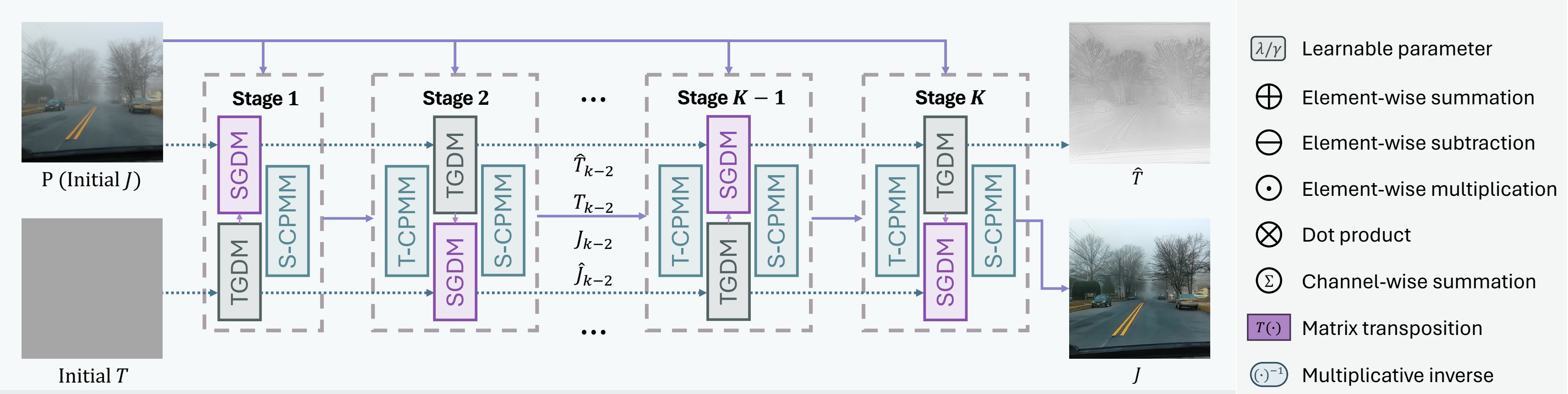
- Same as the proximal mapping process in the transmission optimization, S-CPMM has the similar structure as T-CPMM but different inputs, we denote S-CPMM as $prox_\psi$:

$$\mathbf{J}_k = \mathrm{prox}_\psi(\hat{\mathbf{J}}_k, \hat{\mathbf{T}}_k)$$

- Where the $\hat{\mathrm{J}}_k$ we process by our SGDM can be presented as:

$$\hat{\mathbf{J}}_k = (\hat{\mathbf{T}}_k^\top \hat{\mathbf{T}}_k + \mu_k \mathbf{I})^{-1} \cdot (\hat{\mathbf{T}}_k^\top \mathbf{P} + \hat{\mathbf{T}}_k^\top \hat{\mathbf{T}}_k - \hat{\mathbf{T}}_k^\top + \mu_k \mathbf{J}_{k-1})$$

As the λ_k in transmission optimization, μ_k is also a learnable parameter to bring more generalization capabilities to the network.

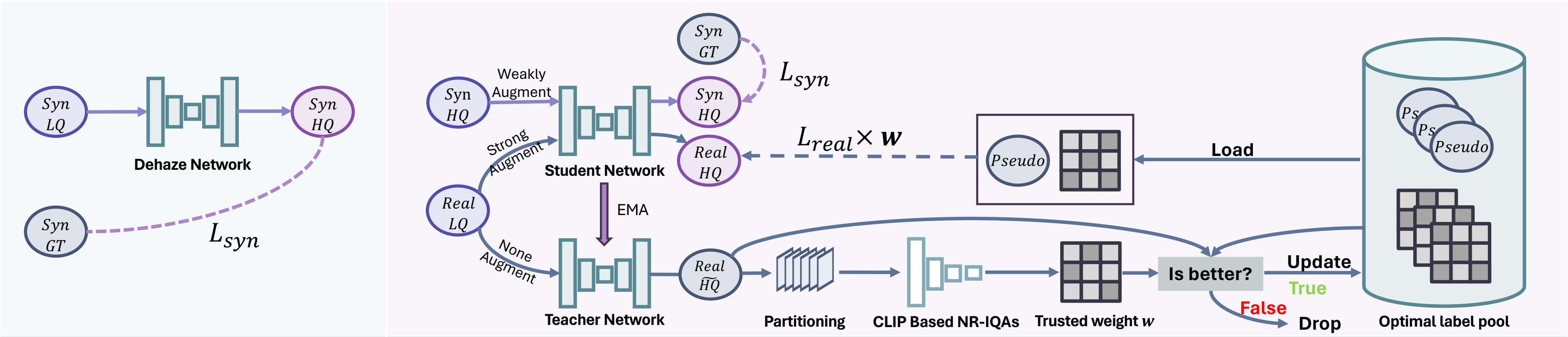


(i) The architecture of the proposed CORUN.

- Each stage of CORUN includes Transmission and Scene Gradient Descent Modules (T&SGDM) paired with Cooperative Proximal Mapping Modules (T&CPMM). **These modules work together to model atmospheric scattering and image scene features, enabling the adaptive capture and restoration of global composite features within the scene.**
- Each CPMM block uses a 4-channel convolution to embed T and J into a feature map. This enables S-CPMM to learn additional scene feature information, such as atmospheric light and blur, assisting SGDM in generating higher-quality dehazed results with more details.
- Our method provides better degradation resistance in the generated results compared to other methods, resulting in higher image quality. It delivers better results in real-world dehazing tasks.

NEURAL INFORMATION PROCESSING SYSTEMS

Coherence-based Pseudo Labeling by Colabator:



(i) Pre-training phase

(ii) Fine-tuning phase with our Colabator

Pre-training network using paired synthetic data.

Fine-tuning network using paired synthetic data and degraded real data by Colabator with only 5000 iter. No additional computational cost during inference.

- **Iterative mean-teacher dehazing:** $\mathbf{P}^R_{\widetilde{HQ}}, \mathbf{T}^R_{\widetilde{HQ}} = f_{\theta_{tea}}(\mathbf{P}^R_{LQ}), \quad \mathbf{P}^R_{\widehat{HQ}}, \mathbf{T}^R_{\widehat{HQ}} = f_{\theta_{stu}}(\mathcal{A}_s(\mathbf{P}^R_{LQ}))$

  This method applies strong data augmentation $\mathcal{A}_s$ to real hazy images, with the teacher network using the original image and the student network using the augmented one, resulting in varying dehazing quality, reducing overfitting and progressively improving supervision reliability.

- **Label trust weighting:** $w = \Psi(\mathrm{norm}(\mathcal{D}(\mathbf{S}^R_{\widetilde{HQ}})) \cdot \mathrm{norm}(\mathcal{Q}(\mathbf{S}^R_{\widetilde{HQ}}))$

  This method assigns reliability weights to locations in pseudo-dehazed images from the teacher network by evaluating haze density and image quality. Using CLIP-based and non-reference metrics ($\mathcal{D}$ and $\mathcal{Q}$), it calculates normalized scores to emphasize clearer, higher-quality regions, improving model supervision.

Coherence-based Pseudo Labeling by Colabator:

- **Optimal label pool:** $\mathbf{P}^R_{HQ}, \mathbf{T}^R_{HQ}, \mathbf{P}^R_{Pse}, \mathbf{T}^R_{Pse}, w_{pse} = \mathcal{C}(\mathbf{P}^R_{LQ}, \theta_{tea}, \theta_{stu}, \mathcal{A}_s, \mathcal{D}(\cdot), \mathcal{Q}(\cdot), \mathcal{P})$

  The optimal label pool $\mathcal{C}$ maintains the best pseudo-labels by updating them only when new dehazed images show improvement, thus stabilizing training and enhancing label reliability within the Colabator framework.

- **Weights update:** $\theta_{tea} = \eta\theta_{tea} + (1-\eta)\theta_{stu}$

  The teacher network updates its weights through an exponential moving average of the student's weights, enabling stable, continuous integration of learned parameters.

Algorithm 1 Optimal label pool process

Require: Haze density evaluator $\mathcal{D}(\cdot)$ and image quality evaluator $\mathcal{Q}(\cdot)$;
Optimal label pool $\mathcal{P}$;
Sample a batch of real hazy images $\{\mathbf{P}^R_{LQ_i}\}_{i=1}^{b}$;
for each $\mathbf{P}^R_{LQ_i}$ **do**
  Get teacher network prediction: $\mathbf{P}^R_{\widetilde{HQ}_i}, \mathbf{T}^R_{\widetilde{HQ}_i} = f_{\theta_{tea}}(\mathbf{P}^R_{LQ_i})$;
  Partition $\mathbf{P}^R_{\widetilde{HQ}_i}$ into $N \times N$ and get $\mathbf{S}^R_{\widetilde{HQ}_i}$;
  Compute score map of $\mathbf{S}^R_{\widetilde{HQ}_i}$: $d_i = \text{norm}(\mathcal{D}(\mathbf{S}^R_{\widetilde{HQ}_i}))$, and $q_i = \text{norm}(\mathcal{Q}(\mathbf{S}^R_{\widetilde{HQ}_i}))$;
  Load $\mathbf{P}^R_{Pse i}, \mathbf{T}^R_{Pse i}, w_{Pse i}, d_{Pse i}, q_{Pse i} = \mathcal{P}(i)$
  if $d_i > d_{Pse i}$ and $q_i > q_{Pse i}$ **then**
    Compute trusted weight: $w_i = \Psi(d_i + q_i)$
    Update $\mathcal{P}(i) = (\mathbf{P}^R_{\widetilde{HQ}_i}, \mathbf{T}^R_{\widetilde{HQ}_i}, w_i, d_i, q_i)$
    Return $\mathbf{P}^R_{\widetilde{HQ}_i}, \mathbf{T}^R_{\widetilde{HQ}_i}, w_i$ as pesudo label.
  else
    Return $\mathbf{P}^R_{Pse i}, \mathbf{T}^R_{Pse i}, w_{pse_i}$ as pesudo label.
  end if
end for

NEURAL INFORMATION PROCESSING SYSTEMS

Results:

Hazy image PDN DAD PSD D4 DGUN RIDCP Ours

DGUN RIDCP Ours

RIDCP Ours Ground truth

Partial Results:

NEURAL INFORMATION PROCESSING SYSTEMS

Metrics	Hazy	PDN [11]	MBDN [14]	DH [15]	DAD [27]	PSD [19]	D4 [26]	RIDCP [7]	DGUN [10]	Ours
FADE↓	2.484	**0.876**	1.363	1.895	1.130	0.920	1.358	0.944	1.111	**0.824**
BRISQUE↓	36.642	30.811	27.672	33.862	32.241	27.713	33.210	**17.293**	27.968	**11.956**
NIMA↑	4.483	4.464	4.529	4.522	4.312	4.598	4.484	**4.965**	4.653	**5.342**

Table 1: Quantitative results on RTTS dataset. **Red** and **blue** indicate the best and the second best.

Class(AP)	Hazy	PDN [11]	MBDN [14]	DH [15]	DAD [27]	PSD [19]	D4 [26]	RIDCP [7]	DGUN [10]	Ours
Bicycle	0.51	0.55	0.54	0.47	0.52	0.52	0.54	**0.57**	0.55	**0.59**
Bus	0.25	0.29	0.27	0.23	0.29	0.25	0.28	**0.32**	**0.31**	**0.31**
Car	0.61	0.65	0.63	0.51	0.65	0.63	0.64	**0.67**	0.66	**0.68**
Motor	0.38	0.45	0.43	0.37	0.38	0.42	0.42	**0.47**	0.46	**0.49**
Person	0.73	**0.76**	0.75	0.69	0.74	0.74	0.75	**0.76**	**0.76**	**0.77**
Mean	0.50	0.54	0.52	0.45	0.52	0.51	0.53	**0.56**	0.55	**0.57**

Table 6: Object detection results on RTTS[40].

Datasets	Metrics	w/o Colabator DGUN	w/ Colabator DGUN	w/o Colabator CORUN	w/ Colabator CORUN (Ours)
RTTS	FADE↓	1.111	0.857	1.091	0.824
	BRISQUE↓	25.085	20.731	16.541	11.956
	NIMA↑	4.813	5.190	4.856	5.342

Table 2: Generalization and Effect of our Colabator.

Datasets	Metrics	w/o Mean-teacher	w/o Trusted weight	w/o Optimal label pool
RTTS	FADE↓	0.912	0.827	0.846
	BRISQUE↓	15.728	16.606	15.707
	NIMA↑	4.921	4.867	5.285

Table 3: Module's Effect of our Colabator.

Datasets	Metrics	Stages			
		1	2	4 (Ours)	6
RTTS	FADE↓	0.785	0.808	0.824	0.839
	BRISQUE↓	15.520	15.151	11.956	16.227
	NIMA↑	5.228	5.281	5.342	5.187

Table 4: Effect of stage number.

Table 11: Ablation of our simplified ASM formula.

ASM formula	NIMA ↑	BRISQUE ↓	FADE↓
w/o simplify	5.203	14.469	0.817
w/ simplify(CORUN+)	5.342	11.956	0.824

Table 13: Effects of integrating our Colabator with more cutting-edge dehazing methods. The gains brought by Colabator are significant.

ASM formula	NIMA ↑	BRISQUE ↓	FADE↓
C2PNet[22]	4.715	34.314	2.064
C2PNet+Colabator	4.823	23.662	1.329
FFA-Net[17]	4.822	33.235	2.080
FFA-Net+Colabator	4.839	29.219	0.958
GDN[16]	5.074	33.051	2.611
GDN+Colabator	5.258	23.691	0.947

Table 8: Ablation of our trusted weights present as a map or value.

Methods	NIMA ↑	BRISQUE ↓	FADE ↓
Only Full	5.229	13.099	0.803
Partition+Full(CORUN+)	5.342	11.956	0.824

Code and Paper

Chengyu's Homepage

NEURAL INFORMATION PROCESSING SYSTEMS

Real-world Image Dehazing with Coherence-based Label Generator and Cooperative Unfolding Network

Chengyu Fang, Chunming He†, Fengyang Xiao, Yulun Zhang†, Longxiang Tang, Yuelin Zhang, Kai Li, Xiu Li†

Thank you for listening !

清華大學 Tsinghua University

Duke UNIVERSITY

上海交通大學 SHANGHAI JIAO TONG UNIVERSITY

中山大學 SUN YAT-SEN UNIVERSITY

香港中文大學 The Chinese University of Hong Kong

Meta