Video Restoration Framework and its Meta-adaptations to Data-poor Conditions

Prashant W Patil[1[0000−0003−2604−6501]], Sunil Gupta[1[0000−0002−3308−1930]], Santu Rana[1[0000−0003−2247−850X]], and Svetha Venkatesh[1[0000−0001−8675−6631]]

A^2I^2, Deakin University, Geelong Warun Ponds Campus, VIC, Australia
prashant.patil@deakin.edu.au

Abstract. Restoration of weather degraded videos is a challenging problem due to diverse weather conditions e.g., rain, haze, snow, *etc.* Existing works handle video restoration for each weather using a different custom-designed architecture. This approach has many limitations. First, a custom-designed architecture for each weather condition requires domain-specific knowledge. Second, disparate network architectures across weather conditions prevent easy knowledge transfer to novel weather conditions where we do not have a lot of data to train a model from scratch. For example, while there is a lot of common knowledge to exploit between the models of different weather conditions at day or night time, it is difficult to do such adaptation. To this end, we propose a generic architecture that is effective for any weather condition due to the ability to extract robust feature maps without any domain-specific knowledge. This is achieved by novel components: spatio-temporal feature modulation, multi-level feature aggregation, and recurrent guidance decoder. Next, we propose a meta-learning based adaptation of our deep architecture to the restoration of videos in data-poor conditions (night-time videos). We show comprehensive results on video de-hazing and de-raining datasets in addition to the meta-learning based adaptation results on night-time video restoration tasks. Our results clearly outperform the state-of-the-art weather degraded video restoration methods. The source code is available at: https://github.com/pwp1208/Meta_Video_Restoration

Keywords: Spatio-temporal Feature Modulation, Meta-adaptation, Day and Night-time Video Restoration.

1 Introduction

Fog, snow, rain, and haze are different types of adverse weather conditions that often degrade the quality of images or videos recorded for computer vision applications such as video surveillance, traffic monitoring, and autonomous driving. These applications routinely involve subtasks such as optical flow estimation [40], object detection [15], depth estimation [33], which use algorithms or models expecting clean data (image or video) as their input. The research on weather-degraded restoration is quite active [22,4,35,17,34]. Earlier works mainly focused

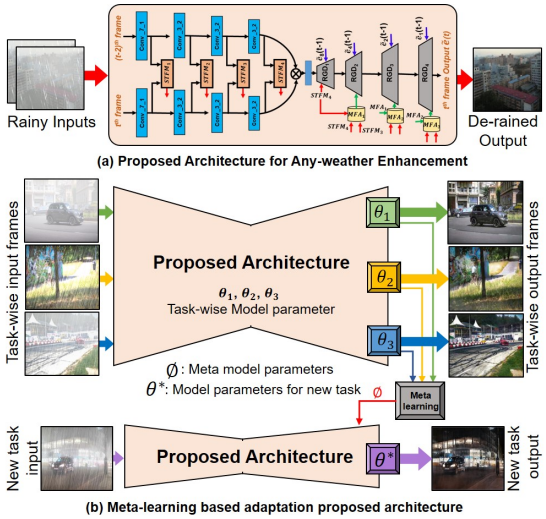

Fig. 1. (a) A generic architecture is provided for restoring any weather degraded video. This architecture outperforms the existing custom-designed architectures and lends itself to easy meta- learning based adaptations permitting sample-efficient learning for novel data-poor weather conditions. (b) A meta-learning (ϕ) adaptation from previous models (θ_1, θ_2, θ_3) to a new model (θ^*) for novel weather condition.

on single image restoration. Most of them are focused on a weather while some of them [22,4] have also extended for multi-weather conditions, *i.e.* use the *same* architecture to train a model for a given weather condition. The extension to video restoration includes [9,44,43,45,49,19]. Many of these models have achieved remarkable success *e.g.,* video de-hazing [18,49], video de-raining [45,44] and video de-raining with veiling effect [43,41].

However, these methods for video restoration have a crucial limitation. They use different architectures for handling different weather conditions. Since a system usually needs to deal with multiple weather conditions, when working with models with disparate architectures, the design process becomes cumbersome. The models of different weather conditions have a great deal of similarity (*e.g.,* in the feature extraction step), which can be exploited to adapt these models to novel conditions that are data-poor using recent machine learning techniques such as few-shot learning [1], meta-learning [11], *etc.* However, this cannot be easily done unless we use the same architecture across different conditions. For example, if using the same architecture, we can use meta-learning to efficiently adapt a *day-time* video restoration model to *night-time* condition. Therefore, we need a generic architecture for video restoration that allows learning across multiple weather conditions and permits transfer to novel conditions.

For designing a single architecture that works across multi-weather video restoration, we need to extract robust feature maps. For this, we propose three modules: *spatio-temporal feature modulator* (STFM) to interlink features at mul-

tiple spatial and temporal scales, *multi-level feature aggregation* module for combining different STFM outputs, and *recurrent guidance decoder* to generate temporally consistent content. This proposed architecture does not require prior domain-specific knowledge and is agnostic to weather conditions. Further, our generic architecture is amenable to model adaptation through meta-learning [11]. We provide an efficient adaptation scheme wherein our architecture can be used in combination with the popular meta-learning algorithm-MAML [11].

In experiments, we first demonstrate that our generic architecture outperforms state-of-the-art (SOTA) for video de-hazing and de-raining tasks using REVIDE [49] and RainSynAll100 [42] datasets respectively. Next, we demonstrate how our model can be efficiently adapted to novel weather conditions. Particularly, we examine adaptation for night-time video restoration that has limited availability of weather degraded night-time videos. In this meta-learning setting, we first train the meta-model and task-wise models for day-time haze, rain, and rain with veiling effects removal tasks. We then show that our meta-model can be adapted to night-time haze, rain, and rain with veiling effects removal in a more sample-efficient manner than training from scratch or fine-tuning the day-time trained models with night-time data. Finally, we also make available a synthetic night-time weather degraded database for de-hazing, de-raining, and de-raining with veiling effects useful to the community given the lack of such datasets. Overview of proposed architecture for weather-degraded restoration and its meta-learning based adaptation for new weather condition is depicted in Fig. 1. Our major contributions are:

- A novel architecture is proposed for any-weather degraded video restoration based on spatio-temporal feature modulation with multi-level feature aggregation and recurrent guidance decoder.
- We propose a meta-learning based adaptation of this architecture for handling novel data-poor weather-degraded conditions.
- We first show comprehensive results on video de-hazing and de-raining datasets. We then show the meta-learning based adaptation results for night-time weather-degraded video restoration. We obtain superior performance for rain, haze, and rain with veiling effect removal for day and night conditions.

To the authors' best knowledge, this is the first video restoration contribution that uses the same architecture across weather conditions, and it is also the first meta-learning adaptation of video restoration models to new conditions. Our approach does not require domain-specific knowledge (*transmission map [41], future frames [44]*) during training, unlike the current SOTA approaches.

2 Literature Survey

Video De-raining Methods: The first study on video rain removal is done in [13], which utilizes the space-time correlation model to capture the dynamics of raindrops. In [24], a hybrid rain model and motion segmentation context information is integrated with a dynamic routing residue recurrent network for

video rain removal. The discriminative prior knowledge-based video rain streak removal approach is proposed by Jiang *et al.* [17]. This approach captures inherent features related to rain streaks based on sparse coding in [20]. Chen *et al.* [5] proposed a novel content alignment and compensation approach for video rain removal. The sparse coding with a multi-scale approach is proposed in [20] to deliver the former characteristic of rain streaks. Any video-based application needs to capture temporal consistency effectively for superior performance. In [44], the temporal correlation and consistency among consecutive video frames are learnt for video rain streak removal. To impose the inter-frame consistency constraint, five successive video frames are used. They estimate the optical flow and warped with input frames. Further, the prediction network is proposed for video rain streak removal. Due to degraded input frames, optical flow may cause many problems. In [41], the robust self-aligned video de-raining approach with transmission depth consistency is proposed. Recently, Yue *et al.* [45] proposed semi-supervised video de-raining approach with dynamic rain generation process.

Video De-hazing Methods: Many algorithms are proposed for image de-hazing with [8,36] and without [9,34] prior information. Dhara [8] *et al.* proposed weighted least squares filtering with adaptive air-light refinement and non-linear color balancing approach for image de-hazing. Zhang *et al.* [48] proposed illumination balancing approach for night-time image de-hazing. Further, maximum reflectance prior [47] and multiple light colors [23] based approaches are proposed for night-time dehazing. These approaches may perform poorly for video de-hazing task and may achieve better results by considering the temporal consistency. In 2018, the first attempt for video de-hazing with multi-frame multi-level fusion strategy was made by Li *et al.* [18]. Further, the transmission map-based video de-hazing module is proposed in [33]. Video de-hazing approaches received less attention as video de-hazing databases were not available. But, a real-world video de-hazing database has become recently available [49].
These existing works are able to handle single weather (haze or rain) efficiently through disparate architectures. Since a system usually needs to deal with multiple weather conditions, when working with models with disparate architectures, the design process becomes cumbersome. Also, using disparate architectures prevents easy adaptation to novel weather-degraded conditions.

Multi-weather Video restoration Methods: Few researchers proposed multi-weather single image restoration approaches [22,46]. First attempt with a gated context aggregation network was proposed by Chen *et al.* [4] to restore the haze and rain-free image directly. In [22], rain, fog, snow, and adherent raindrops weather conditions are handled using multiple task-specific encoders with neural architectures. Despite excellent performance in multi-weather degraded image restoration, these methods may fail for video restoration due to a lack of temporal consistency. In this context, Yang *et al.* [42] proposed the first video de-raining approach by considering the veiling effect. The veiling effect is rain-streak accumulation in the line of sight. Similar to [44], the recurrent multi-frame de-raining with veiling effect approach is proposed in [43] with physics model and adversarial learning. To maintain the temporal consistency, they considered

five consecutive frames including **future frames** to enhance the current frame with multi-stage process. Recently, Li *et al.* [19] proposed multi-frame based rain and snow removal approach.

Meta-learning Methods: Meta-learning aims to extract meta-knowledge from historical tasks to accelerate learning on new tasks by transferring previously learned knowledge. Meta-learning has been applied to many applications like video interpolation [6], object segmentation [2], action recognition [7] and object tracking [38]. In [6], authors considered the interpolation of a single video as one task. They analysed the effect of existing SOTA networks with scratch training, fine-tuning and meta-learning. Xinjian *et al.* [12] used meta-learning to transfer embeddings across rainy and clean images. In contrast, meta-learning in our architecture is used to adapt the day time (haze, rain and rain with veil) video restoration model to night-time conditions. However, none of these works have focused on multi-weather degradation restoration.

3 Proposed Video restoration Framework

In this section, we describe the proposed architecture and its meta-learning adaptation to adapt the model for data-poor weather conditions.

3.1 Network Architecture

The overview of the proposed architecture for multi-weather video restoration is illustrated in Fig. 2 and 3 which comprises of three major components: (1) Spatio-temporal feature modulation (STFM), (2) Multi-level feature aggregation (MFA), and (3) Recurrent guided decoder (RGD).

Spatio-temporal Feature Modulation: The efficient feature fusion of different time instances plays an important role in video processing applications. To effectively interlink the multi-frame features, we propose the STFM module. The STFM module effectively extracts the multi-frame features through scale and temporal modulations. As $(t-1)^{th}$ frame output is provided at decoder recurrently, the t^{th} and $(t-2)^{th}$ frames are given to two different encoder paths as inputs. The feature maps of both the frames at each encoder level are given to STFM modules. In the STFM module, each of these feature maps (t^{th} and $(t-2)^{th}$ frames) are passed through multi-scale convolution block (convolution with filter size $1, 3, 5$). Further, these multi-scale feature maps are processed through scale modulation to interlink the information at different scales. The scale modulation helps the network to boost the feature maps by incorporating the feature maps from multi-scales. Scale modulation (SM) is defined as:

$$SM_{a,b} = \alpha\,[f_{Sc\_a} \copyright f_{Sc\_b}] + \beta; \quad a \neq b \tag{1}$$

where, f_{Sc_a}, f_{Sc_b} are outcomes of multi-scale convolution block ($a, b \in (1, 3, 5)$), $\alpha = \gamma[\mathbb{C}^3\{\underset{c\in(1,C)}{\text{avg}}\,(f_{Sc\_a}, f_{Sc\_b})\}]$, $\beta = \gamma[\mathbb{C}^3\{f_{Sc\_a} \copyright f_{Sc\_b}\}]$, γ is global average

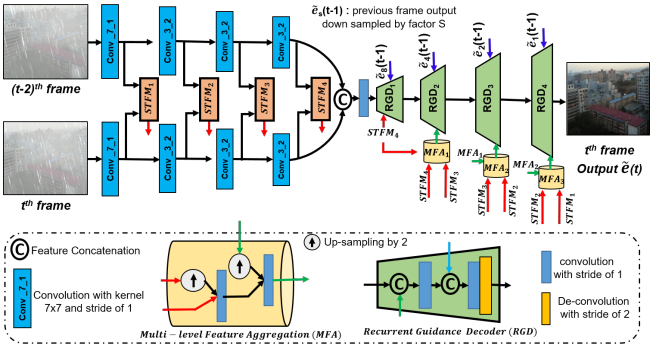

Fig. 2. Proposed architecture for weather-degraded video restoration (*STFM: spatio-temporal feature modulation*).

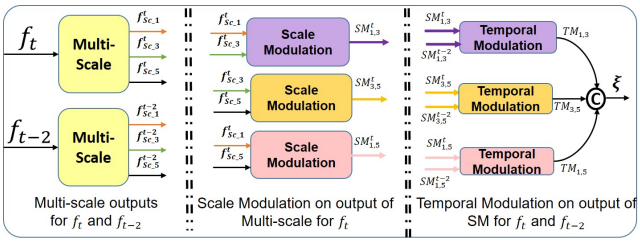

Fig. 3. Proposed spatio-temporal feature modulation (STFM).

pooling, ⓒ is concatenation, C is total number of channels and $\mathbb{C}^3$ is convolution with kernel 3×3 (see Fig. 3). As video-based frameworks need to deal with multi-frame information correlation, the scale modulated feature maps are temporally modulated. Temporal features of t^{th} and $(t-2)^{th}$ frames are modulated as:

$$TM_{a,b} = \alpha\left[SM^{t}_{a,b} \text{ⓒ} SM^{t-2}_{a,b}\right] + \beta; \quad a \neq b \tag{2}$$

where, $\alpha = \gamma[\mathbb{C}^3\{\underset{c\in(1,C)}{\text{avg}}(SM^{t}_{a,b}, SM^{t-2}_{a,b})\}]$, and $\beta = \gamma[\mathbb{C}^3\{SM^{t}_{a,b} \text{ⓒ} SM^{t-2}_{a,b}\}]$. This process helps to learn the inter-frame information. Here, the α and β in both SM and TM are mainly proposed for scaling and shifting (modulation) of the multi-scale feature maps and the feature maps of different time instance respectively. Finally, three temporally modulated features ($TM_{1,3}$, $TM_{3,5}$, $TM_{1,5}$) are merged and considered as the output (ξ) of STFM module.

Multi-level Feature Aggregation: The proposed STFM module interlinks the multi-frame feature maps through scale and temporal modulation. Multi-level feature interlinking is a crucial task to enlarge the overall receptive field of the network. To do this, the multi-level feature aggregation (MFA) module is proposed. This MFA module is hierarchical, which takes the multi-level STFM

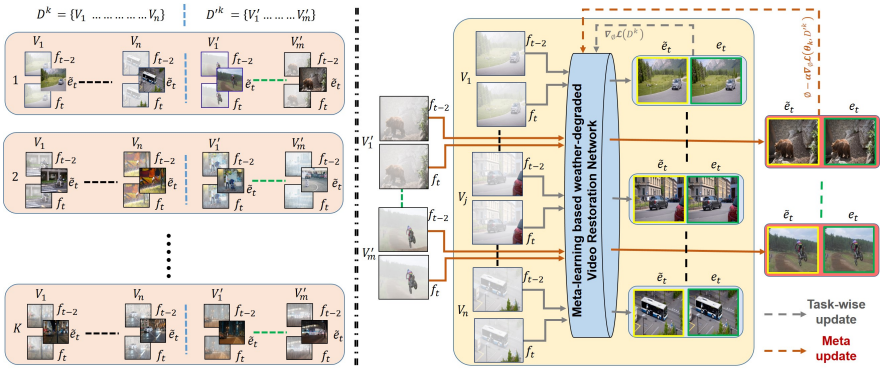

Fig. 4. Overview of the proposed meta-learning based weather-degraded video restoration framework. ***Left:*** Each task $\in (1, K)$ consists of n number of training videos and m number of validation videos. These train videos, V_1 to V_n, are used for task-wise update (*i.e.,* the inner loop) and validation videos, V_1' to V_m', are used for meta-update (*i.e.,* the outer loop). ***Right:*** The training videos V_1 to V_n from each task 1 to k in D^k are used to adapt θ_k using task-wise optimizers for inner loop and validation videos V_1' to V_m' from each task 1 to k in D'^k are used to adapt meta optimizer for outer loop.

features to capture and fuse the effective information at different levels. Also, compared to normal feature aggregation/concatenation operation, the proposed MFA obtains a wide-ranging vision field to further guide the feature maps as:

$$MFA_l = \begin{cases} \left[\xi_{(4-l)} \copyright \xi^2_{(5-l)}\right] \copyright \xi^2_{(5-l)}; \ l = 1 \\ \left[xi_{(4-l)} \copyright xi^2_{(5-l)}\right] \copyright MFA^2_{(l-1)}; \ l > 1 \end{cases} \tag{3}$$

where, ξ are STFM feature maps, ξ^2 are corresponding up-sampled STFM feature maps and MFA^2 are respective up-sampled MFA feature maps. In the proposed network, we have used three MFA blocks with sequential input.

Recurrent Guided Decoder: The proposed RGD module takes advantage of the spatio-temporally modulated features, multi-level feature aggregation features, and previous frame output feedback for the effective restoration of the current frame. The RGD module at each scale correlates the previous frame output maps ($\tilde{e}(t-1)$) and respective MFA feature maps to learn the temporally consistant content of the current frame. Initially, the MFA feature maps are merged with previous scale RGD feature maps. These features are correlated through the convolution operation. In [28], the authors argued that the feedback from either decoder level features or simply the output of the previous frame works very effectively for video-based applications. Therefore, these correlated feature maps are merged with the subsequent scale of the previous frame output maps ($\tilde{e}_s(t-1)$) to restore temporally consistent current frame.

3.2 Learning the Model Parameters

The proposed network parameters are optimized using $\mathbb{L}_1$ *i.e.* Least absolute deviations as:

$$\mathbb{L}_1 = |(e_t - \tilde{e_t})| \tag{4}$$

where, e_t is target frame and $\tilde{e}_t$ is an restored frame using proposed architecture. We note that the $\tilde{e}_t$ is a function of f_t, f_{t-2} and $\tilde{e}_{t-1}$. Along with $\mathbb{L}_1$, to guide the model with textural and structural information, the perceptual loss with pre-trained VGG19 model [37] is calculated as:

$$\mathbb{L}_P = \sum_{l=1}^{L} \|\psi_l(\tilde{e}_t) - \psi_l(e_t)\|_1 \tag{5}$$

where, $\psi_l(.)$ represents l^{th} pooling layer of VGG-19 model. Further, we have considered structural similarity index (SSIM) loss to preserve high frequency information [31]. The SSIM loss function is defined as:

$$\mathbb{L}_S = 1 - SSIM(\tilde{e}_t, e_t) \tag{6}$$

The combination of SSIM loss with edge loss has been shown to work well [10]. So, the edge loss is also considered to focus on the edge restoration while training the proposed network. Edge loss is formulated with Sobel operator ($\mathbb{S}$) as:

$$\mathbb{L}_{ed} = \|\mathbb{S}(\tilde{e}_t) - \mathbb{S}(e_t)\|_1 \tag{7}$$

Thus, the overall loss ($\mathbb{L}_{Total}$) for training the proposed network is given as:

$$\mathbb{L}_{Total} = \lambda_1\mathbb{L}_1 + \lambda_{ed}\mathbb{L}_{ed} + \lambda_P\mathbb{L}_P + \lambda_S\mathbb{L}_S \tag{8}$$

where, λ_{loss} are the weights assigned for the respective loss functions.

3.3 Meta-learning based Adaptation

The proposed framework deals with six different weather-degraded scenarios: day-time (haze, rain, rain with veiling effect), and night-time (haze, rain and rain with veiling effect). The overview of the proposed meta-learning framework is depicted in Fig. 4 and summarised in Algorithm 1. In general, we have K different tasks in training set, $D = \left\{D^1, D^2, ...D^K\right\}$ and dataset for each task contains n data points as $D^k = \{V_1, V_2, ...V_n\}$ where $V_n = \{(f_{t-2}, f_t), e_t \mid t > 2\}$ as pair of inputs (f_{t-2}, f_t) and target frame (e_t) at time $t \in (3, ...F)$ and F being the total number of frames in each video. Similarly, the validation set used to update the meta model $D' = \left\{D'^1, D'^2, ...D'^K\right\}$ contains $D'^k = \{V'_1, V'_2, ...V'_m\}$. ***Left*** part of Fig. 4 shows the training and validation splits.

While training, a copy of the proposed architecture is kept as the meta-model and denoted as ϕ. The proposed architecture is first trained task-wise

Algorithm 1 : Pseudocode for meta-adaptation

Input: Training D^k and validation D'^k datasets, learning rate α
Initialize ϕ
while not done **do**
  for k in $\{1, 2, ...K\}$ **do**
    Initialize θ_k
    Optimize the θ_k as
    $\theta_k \leftarrow \phi - \alpha \bigtriangledown_\phi \mathbb{L}_{Total}\left(\theta_k, D^k\right)$
  end for
  Update $\phi \leftarrow \phi - \alpha \sum_{k=1}^{K} \bigtriangledown_\phi \mathbb{L}_{Total}\left(\theta_k, D'^k\right)$
end while

using training data D to learn task-specific parameters θ_k. The task-specific parameters are optimized using D^k as:

$$\theta_k \leftarrow \phi - \alpha \bigtriangledown_\phi \mathbb{L}_{Total}\left(\theta_k, D^k\right) \tag{9}$$

where, ϕ is meta-model parameter, α is learning rate, θ_k are task-wise parameters to be optimized using gradients $\bigtriangledown_\phi$ with respective losses ($\mathbb{L}_{Total}$) on training split (D^k) and by $\mathbb{L}_{Total}(\theta_k, D^k)$, we mean the $\mathbb{L}_{Total}$ computed on model with parameters θ_k using dataset D^k. The gradients of task-wise adapted model are then used to update the meta model parameters as:

$$\phi \leftarrow \phi - \alpha \sum_{k=1}^{K} \bigtriangledown_\phi \mathbb{L}_{Total}\left(\theta_k, D'^k\right) \tag{10}$$

4 Multi-weather Database Generation

The collection of real-world weather degraded day-night video with respective clean video is a challenging task. Therefore, many weather-specific synthetic video databases are introduced in the literature only with day-time scenarios. No single video dataset is generated for night-time weather degraded restoration. Therefore, in this work, we synthetically generate the day and night-time haze, rain and rain with veiling effect video datasets for meta-learning based training and new task adaptation purpose. For the synthetic day-time multi-weather video database generation, we have used a popular outdoor DAVIS-2016 [29] video database. Also, the night-time videos are downloaded from https://www.pexels.com/videos/. The depth maps from [32] and procedure for synthetic database generation is adapted from [21]. Few samples from the synthesized day and night-time haze, rain and rain with veiling effect datasets are shown in Fig. 5. In total, 30 (20: task-wise training *i.e. task-specific model update* and 10: validation *i.e. meta model update*) videos for each day-time and 30 (10: meta-training and 20: meta-model testing) videos for each night-time weather degraded tasks are generated for meta-learning based weather-degraded video restoration (*see supp. material for more details*).

Fig. 5. Synthetically generated video frames (*first three columns: day-time and last three columns: night-time*).

Table 1. Quantitative results with GDN [26], DuRN [27], KDNN [14], FFA [30], EDVR [39], MSBD [9], IDN [49] on REVIDE [49] database for video de-hazing (*all quantitative values are collected from [49]* and PM: Proposed Method).

Methods →	GDN	DuRN	KDNN	FFA	EDVR	MSBD	IDN	PM
PSNR	19.69	18.51	16.32	16.65	21.22	22.01	23.21	**26.36**
SSIM	0.8545	0.8272	0.7731	0.8133	0.8707	0.8759	0.8836	**0.9044**

5 Experiments

5.1 Implementation Details

Losses in Eq. (8) are considered when training a proposed architecture for de-hazing and de-raining and adapting it to night-time weather conditions. The values of λ_{loss} are set (verified experimentally) as $\lambda_{ed} = 0.7$, $\lambda_P = 0.5$, $\lambda_S = 0.5$ and $\lambda_1 = 1$. The proposed network implemented with Tensorflow 2.0 library on NVIDIA DGX Tesla V100 32 GB GPU.

We compare the proposed method with existing SOTA approaches on REVIDE and RainSynAll100 datasets for video de-hazing and de-raining tasks. We used the same train-test splits as provided in the respective datasets. Weights are initialized randomly and optimized with Adam optimizer. For the meta-learning based adaptation results, we use our synthetically generated day and night-time haze, rain, and rain with veiling effect video datasets: training the model on day videos and then efficiently adapting to night conditions. We compare the results of the meta-learnt model vs. learning from scratch on night conditions. PSNR (luminance channel) and SSIM are used as evaluation metrics.

5.2 Analysis of Proposed Architecture

De-hazing: The proposed network is tested on REVIDE [49] video de-hazing database. The SOTA de-hazing methods GDN [26], DuRN [27], KDNN [14], MSBDN [9], FFA [30], EDVR [39], CG-IDN [49] are used for analysis. The quantitative results' analysis of proposed method with SOTA methods is given in Table 1 for video de-hazing. The qualitative results on REVIDE database and real world video are shown in Fig. 6. From Table 1, and Fig. 6, we can see that the proposed network achieves superior performance for video de-hazing. We

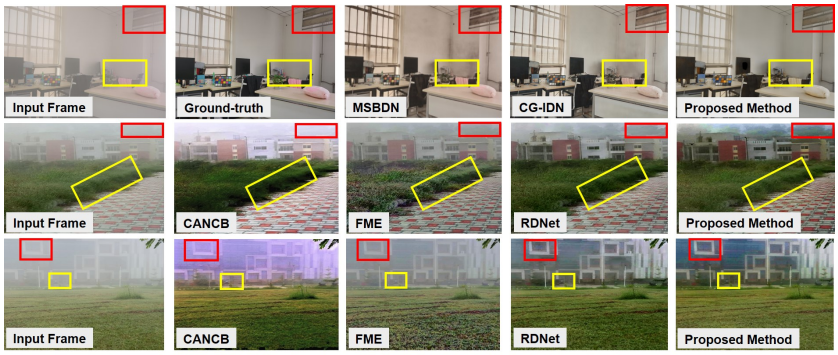

Fig. 6. Qualitative result analysis (*first row: REVIDE, last two rows: Real world*) for video de-hazing (MSBDN [9], CG-IDN [49] and CANCB [8], FME [51], RDNet [50]).

Table 2. Quantitative analysis with DIP [16], EVD [18], SCNN [5], MRF [3], J4RNet [25], DLF [42] and RMFD [43] on RainSynAll100 [42] database († *and* ‡ *represent method is used as pre and post-processing respectively*).

Methods →	DIP† +EVD‡	SCNN† +EVD‡	DIP† + MRF‡	SCNN† +MRF‡	J4RNet	DLF	RMFD	PM
PSNR	18.28	17.87	18.79	18.39	22.93	25.72	25.14	**28.39**
SSIM	0.6804	0.6423	0.6914	0.6469	0.7746	0.8989	0.9172	**0.9317**

note that CG-IDN [49] has ∼**23M** compared our proposed network with ∼**10M** parameters. *Computational complexity analysis is given in supp. material.*

De-raining with Veiling Effect: The proposed network is tested on RainSynAll100 [42] dataset. J4RNet [25], DLF[42], RMFD [43] SOTA methods are used as baselines. To do more analysis, the combination of de-hazing and de-raining approaches are used similar to [43]. The quantitative result analysis for RainSynAll100 datasets is given in the Table 2. Also, the proposed architecture compared qualitatively on RainSynAll100 database and real world rainy videos is illustrated in Fig. 7 (*see Section 3 from supp. material*). 3dB performance improvement is achieved compared to recent RMFD [43] on RainSynAll100 dataset. Also from visual results, it is evident that the DLF and RFMD suffers from rainy streaks effect, veil effect and true color restoration whereas the proposed architecture produces the results without any rain streaks with veil effect removal and true color restoration. Also, the RMFD [43] has ∼**29M** whereas our proposed network has only ∼**10M** parameters.

5.3 Ablation Study on Proposed Architecture

The REVIDE database is used to examine the individual contributions (STFM, MFA, $\tilde{e}_{t-1}$) in the proposed network in terms of average PSNR and SSIM.

Fig. 7. Qualitative result analysis (*first row: RainSynAll100* and *last two rows: Real world videos*) for video de-raining with veiling effect (DLF [42] and RMFD [43]).

Table 3. Ablation study analysis of proposed modules (M: Modulation, MFA: multi-level feature aggregation, and $\tilde{e}_{t-1}$: feedback of previous frame output).

Network ↓	Scale M	Temporal M	MFA	$\tilde{e}_{t-1}$	PSNR	SSIM
I					22.74	0.8489
II	✓				23.69	0.8566
III		✓			23.95	0.8609
IV	✓	✓			24.64	0.8749
V	✓	✓	✓		25.42	0.8814
VI	✓	✓	✓	✓	**26.36**	**0.9044**

The STFM (comprising of SM and TM) interlinks the multi-frame features spatially and temporally. **How does this scale and temporal feature modulation help the network to integrate the effective multi-frame features?** To scrutinize this, the results of the proposed network are analyzed with and without scale and temporal modulation by keeping all other modules the same and results are given in Table 3. STFM module helps the network for inter-frame feature fusion which yields towards effective restoration. This is easily conveyed from results reported in Table 3 (Networks II - IV).

Next, we study the multi-level feature aggregation - the MFA module. **Is interlinking multi-level feature maps through MFA effective?** To analyse this, we examine the accuracy of the proposed network with and without the MFA module. From Network IV and V of Table 3, it is clear that the presence of the proposed MFA module improves performance.

As the motion between two consecutive frames is very minute, the t and $(t-2)$ frames are given as input to get temporal consistency. The enhanced frame $\tilde{e}_{t-1}$ is used to further ensure temporal consistency. **Whether sharing of the previous enhanced frame helps the network to get temporally consistent results?** We analyse the efficiency of the network with and without

Table 4. Analysis of scratch (Scrt), Fine, Combined (Comb) and meta training for night-time video restoration tasks.

Metrics →	PSNR				SSIM			
Night Tasks ↓	Scrt	Fine	Comb	Meta	Scrt	Fine	Comb	Meta
De-hazing	21.99	22.01	22.97	**23.67**	0.6505	0.6615	0.6995	**0.7178**
De-raining	23.57	22.78	23.45	**24.71**	0.6695	**0.7102**	0.6845	0.7029
De-raining w/ veil	22.21	22.88	22.69	**23.49**	0.6723	0.6845	0.6937	**0.7171**

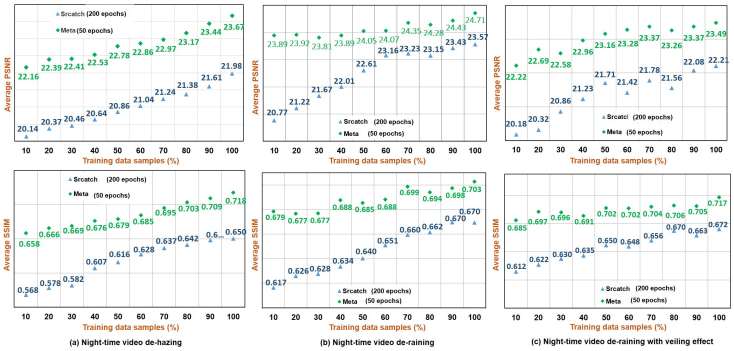

Fig. 8. Scratch and meta-training analysis with respect to different training samples in terms of average PSNR and SSIM for night-time weather degraded restoration tasks.

this recurrent guidance. From results reported in Table 3 (Networks VI), it is clear that conditioning the decoder with $(t-1)^{th}$ frame output leads to better temporal consistency. Also, we use 2 past frames and argue that our method will be less affected by any sudden temporal change (e.g. the first few frames are bright while the later ones are dark) compared to a method [43] that uses more number of previous frames as our restoration is less reliant on past frames. *The ablation study on losses is provided in the supp. material.*

5.4 Analysis of Meta-Adaptation

The meta-learning based adaptation of the proposed architecture is trained end-to-end. Initially, the task-wise and meta-model are initialised randomly and follow the iterative steps for inner and outer loop parameter optimization with day-time tasks (*haze, rain, and rain with veiling effects*) following Algorithm 1. After 250 epochs, the trained meta-model is used to adapt to new tasks: night-time (*haze, rain and rain with veiling effects*) weather degraded video restoration.

In this section, the proposed network is analysed with scratch training, fine-tuning, combined training and meta-learning based adaptation for night-time practical scenarios like haze, rain, and rain with veiling effect. As meta-learning helps the proposed architecture for quick adaptation of new task with few numbers of training samples, we compared the effectiveness of meta-learning based

Fig. 9. Qualitative analysis (1^{st} row: synthetic and 2^{nd} row: real world data) with scratch, fine-tuning and meta-training for night-time de-raining with veiling effect.

adaptation with scratch training in terms of the number of training data samples on night-time video restoration tasks. The efficiency of meta learning in term of average PSNR and SSIM is depicted in Fig. 8 for night-time video dehazing, de-raining and de-raining with veiling effect tasks. It shows that meta-adaptation with just 10% of total training samples already outperforms the model trained from scratch on all 100% of the training samples. This analysis shows the quick adaptation ability of the proposed architecture with few samples using the meta-learning approach. Using 100% training samples in the three cases (scratch training, fine-tuning, and meta-learning), Table 4 compares the performance and shows that the meta adaptation performs the best (see Fig. 9).

6 Conclusion

This work is a general restoration framework which benefits the day, and night-time weather degraded video restoration through the proposed architecture and its meta-learning based adaptation. The proposed architecture makes use of multi-frame based spatio-temporal feature modulation with multi-level feature aggregation and recurrent guidance decoder. Further, the proposed work incorporates the meta-learning based adaptation of the proposed architecture for weather degraded night-time video restoration. To the best of the authors' knowledge, this is the first video restoration attempt to address the problem caused by diverse weather (haze, rain, and rain with veiling effect) in day and night-time through meta-learning based adaptation. As night-time weather degraded video restoration receives less attention due to the limited availability of datasets, we provided the synthetic comprehensive night-time haze, rain, and rain with veiling effects datasets. The comprehensive results on video de-hazing and de-raining datasets in addition to the meta-learning based adaptation on night-time weather degraded video restoration proves the effectiveness of the proposed architecture.

Acknowledgement

This research was partially funded by the Australian Government through the Australian Research Council (ARC). Prof. Svetha Venkatesh is the recipient of an ARC Australian Laureate Fellowship (FL170100006).

References

1. Baik, S., Choi, J., Kim, H., Cho, D., Min, J., Lee, K.M.: Meta-learning with task-adaptive loss function for few-shot learning. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). pp. 9465–9474 (October 2021) 2
2. Behl, H.S., Naja, M., Arnab, A., Torr, P.H.: Meta-learning deep visual words for fast video object segmentation. In: 2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). pp. 8484–8491. IEEE (2020) 5
3. Cai, B., Xu, X., Tao, D.: Real-time video dehazing based on spatio-temporal mrf. In: Pacific Rim conference on multimedia. pp. 315–325. Springer (2016) 11
4. Chen, D., He, M., Fan, Q., Liao, J., Zhang, L., Hou, D., Yuan, L., Hua, G.: Gated context aggregation network for image dehazing and deraining. In: 2019 IEEE winter conference on applications of computer vision (WACV). pp. 1375–1383. IEEE (2019) 1, 2, 4
5. Chen, J., Tan, C.H., Hou, J., Chau, L.P., Li, H.: Robust video content alignment and compensation for rain removal in a cnn framework. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 6286–6295 (2018) 4, 11
6. Choi, M., Choi, J., Baik, S., Kim, T.H., Lee, K.M.: Scene-adaptive video frame interpolation via meta-learning. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 9444–9453 (2020) 5
7. Coskun, H., Zia, M.Z., Tekin, B., Bogo, F., Navab, N., Tombari, F., Sawhney, H.: Domain-specific priors and meta learning for few-shot first-person action recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence pp. 1–1 (2021). https://doi.org/10.1109/TPAMI.2021.3058606 5
8. Dhara, S.K., Roy, M., Sen, D., Biswas, P.K.: Color cast dependent image dehazing via adaptive airlight refinement and non-linear color balancing. IEEE Transactions on Circuits and Systems for Video Technology **31**(5), 2076–2081 (2020) 4, 11
9. Dong, H., Pan, J., Xiang, L., Hu, Z., Zhang, X., Wang, F., Yang, M.H.: Multi-scale boosted dehazing network with dense feature fusion. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 2157–2167 (2020) 2, 4, 10, 11
10. Dudhane, A., Biradar, K.M., Patil, P.W., Hambarde, P., Murala, S.: Varicolored image de-hazing. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 4564–4573 (2020) 8
11. Finn, C., Abbeel, P., Levine, S.: Model-agnostic meta-learning for fast adaptation of deep networks. In: International Conference on Machine Learning. pp. 1126–1135. PMLR (2017) 2, 3
12. Gao, X., Wang, Y., Cheng, J., Xu, M., Wang, M.: Meta-learning based relation and representation learning networks for single-image deraining. Pattern Recognition **120**, 108124 (2021) 5
13. Garg, K., Nayar, S.K.: Detection and removal of rain from videos. In: Proceedings of the 2004 IEEE Computer Society Conference on CVPR, 2004. CVPR 2004. vol. 1, pp. I–I. IEEE (2004) 3
14. Hong, M., Xie, Y., Li, C., Qu, Y.: Distilling image dehazing with heterogeneous task imitation. In: Proceedings of the IEEE/CVF Conference on CVPR. pp. 3462–3471 (2020) 10
15. Huang, Z., Zou, Y., Kumar, B., Huang, D.: Comprehensive attention self-distillation for weakly-supervised object detection. Advances in Neural Information Processing Systems **33** (2020) 1

16. Jiang, T.X., Huang, T.Z., Zhao, X.L., Deng, L.J., Wang, Y.: A novel tensor-based video rain streaks removal approach via utilizing discriminatively intrinsic priors. In: Proceedings of the ieee conference on CVPR. pp. 4057–4066 (2017) 11
17. Jiang, T.X., Huang, T.Z., Zhao, X.L., Deng, L.J., Wang, Y.: Fastderain: A novel video rain streak removal method using directional gradient priors. IEEE Transactions on Image Processing **28**(4), 2089–2102 (2018) 1, 4
18. Li, B., Peng, X., Wang, Z., Xu, J., Feng, D.: End-to-end united video dehazing and detection. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 32 (2018) 2, 4, 11
19. Li, M., Cao, X., Zhao, Q., Zhang, L., Meng, D.: Online rain/snow removal from surveillance videos. IEEE Transactions on Image Processing **30**, 2029–2044 (2021) 2, 5
20. Li, M., Xie, Q., Zhao, Q., Wei, W., Gu, S., Tao, J., Meng, D.: Video rain streak removal by multiscale convolutional sparse coding. In: Proceedings of the IEEE conference on CVPR. pp. 6644–6653 (2018) 4
21. Li, R., Cheong, L.F., Tan, R.T.: Heavy rain image restoration: Integrating physics model and conditional adversarial learning. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 1633–1642 (2019) 9
22. Li, R., Tan, R.T., Cheong, L.F.: All in one bad weather removal using architectural search. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 3175–3185 (2020) 1, 2, 4
23. Li, Y., Tan, R.T., Brown, M.S.: Nighttime haze removal with glow and multiple light colors. In: Proceedings of the IEEE international conference on computer vision. pp. 226–234 (2015) 4
24. Liu, J., Yang, W., Yang, S., Guo, Z.: D3R-Net: Dynamic routing residue recurrent network for video rain removal. IEEE Transactions on Image Processing **28**(2), 699–712 (2018) 3
25. Liu, J., Yang, W., Yang, S., Guo, Z.: Erase or fill? deep joint recurrent rain removal and reconstruction in videos. In: Proceedings of the IEEE conference on CVPR. pp. 3233–3242 (2018) 11
26. Liu, X., Ma, Y., Shi, Z., Chen, J.: Griddehazenet: Attention-based multi-scale network for image dehazing. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 7314–7323 (2019) 10
27. Liu, X., Suganuma, M., Sun, Z., Okatani, T.: Dual residual networks leveraging the potential of paired operations for image restoration. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 7007–7016 (2019) 10
28. Patil, P.W., Biradar, K.M., Dudhane, A., Murala, S.: An end-to-end edge aggregation network for moving object segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 8149–8158 (2020) 7
29. Perazzi, F., Pont-Tuset, J., McWilliams, B., Van Gool, L., Gross, M., Sorkine-Hornung, A.: A benchmark dataset and evaluation methodology for video object segmentation. In: Proceedings of the IEEE conference on CVPR. pp. 724–732 (2016) 9
30. Qin, X., Wang, Z., Bai, Y., Xie, X., Jia, H.: FFA-Net: Feature fusion attention network for single image dehazing. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 34, pp. 11908–11915 (2020) 10
31. Que, Y., Li, S., Lee, H.J.: Attentive composite residual network for robust rain removal from single images. IEEE Transactions on Multimedia (2020) 8

32. Ranftl, R., Lasinger, K., Hafner, D., Schindler, K., Koltun, V.: Towards robust monocular depth estimation: Mixing datasets for zero-shot cross-dataset transfer. IEEE Transactions on Pattern Analysis and Machine Intelligence (2020) 9
33. Ren, W., Zhang, J., Xu, X., Ma, L., Cao, X., Meng, G., Liu, W.: Deep video dehazing with semantic segmentation. IEEE Transactions on Image Processing **28**(4), 1895–1908 (2018) 1, 4
34. Shao, Y., Li, L., Ren, W., Gao, C., Sang, N.: Domain adaptation for image dehazing. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 2808–2817 (2020) 1, 4
35. Sharma, A., Tan, R.T.: Nighttime visibility enhancement by increasing the dynamic range and suppression of light effects. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 11977–11986 (2021) 1
36. Shin, J., Kim, M., Paik, J., Lee, S.: Radiance–reflectance combined optimization and structure-guided norm for single image dehazing. IEEE Transactions on Multimedia **22**(1), 30–44 (2019) 4
37. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556 (2014) 8
38. Wang, G., Luo, C., Sun, X., Xiong, Z., Zeng, W.: Tracking by instance detection: A meta-learning approach. In: Proceedings of the IEEE/CVF conference on Computer Vision and Pattern Recognition. pp. 6288–6297 (2020) 5
39. Wang, X., Chan, K.C., Yu, K., Dong, C., Change Loy, C.: Edvr: Video restoration with enhanced deformable convolutional networks. In: Proceedings of the IEEE/CVF Conference on CVPR Workshops. pp. 0–0 (2019) 10
40. Yan, W., Sharma, A., Tan, R.T.: Optical flow in dense foggy scenes using semi-supervised learning. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 13259–13268 (2020) 1
41. Yan, W., Tan, R.T., Yang, W., Dai, D.: Self-aligned video deraining with transmission-depth consistency. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 11966–11976 (2021) 2, 3, 4
42. Yang, W., Liu, J., Feng, J.: Frame-consistent recurrent video deraining with dual-level flow. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 1661–1670 (2019) 3, 4, 11, 12
43. Yang, W., Tan, R.T., Feng, J., Wang, S., Cheng, B., Liu, J.: Recurrent multi-frame deraining: Combining physics guidance and adversarial learning. IEEE Transactions on Pattern Analysis and Machine Intelligence (2021) 2, 4, 11, 12, 13
44. Yang, W., Tan, R.T., Wang, S., Liu, J.: Self-learning video rain streak removal: When cyclic consistency meets temporal correspondence. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 1720–1729 (2020) 2, 3, 4
45. Yue, Z., Xie, J., Zhao, Q., Meng, D.: Semi-supervised video deraining with dynamical rain generator. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 642–652 (2021) 2, 4
46. Zamir, S.W., Arora, A., Khan, S., Hayat, M., Khan, F.S., Yang, M.H., Shao, L.: Multi-stage progressive image restoration. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 14821–14831 (2021) 4
47. Zhang, J., Cao, Y., Fang, S., Kang, Y., Wen Chen, C.: Fast haze removal for nighttime image using maximum reflectance prior. In: Proceedings of the IEEE conference on Computer Vision and Pattern Recognition. pp. 7418–7426 (2017) 4

48. Zhang, J., Cao, Y., Wang, Z.: Nighttime haze removal based on a new imaging model. In: 2014 IEEE International Conference on Image Processing (ICIP). pp. 4557–4561. IEEE (2014) 4
49. Zhang, X., Dong, H., Pan, J., Zhu, C., Tai, Y., Wang, C., Li, J., Huang, F., Wang, F.: Learning to restore hazy video: A new real-world dataset and a new method. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 9239–9248 (2021) 2, 3, 4, 10, 11
50. Zhao, S., Zhang, L., Shen, Y., Zhou, Y.: RefineDNet: A weakly supervised refinement framework for single image dehazing. IEEE Transactions on Image Processing **30**, 3391–3404 (2021) 11
51. Zhu, Z., Wei, H., Hu, G., Li, Y., Qi, G., Mazur, N.: A novel fast single image dehazing algorithm based on artificial multiexposure image fusion. IEEE Transactions on Instrumentation and Measurement **70**, 1–23 (2020) 11