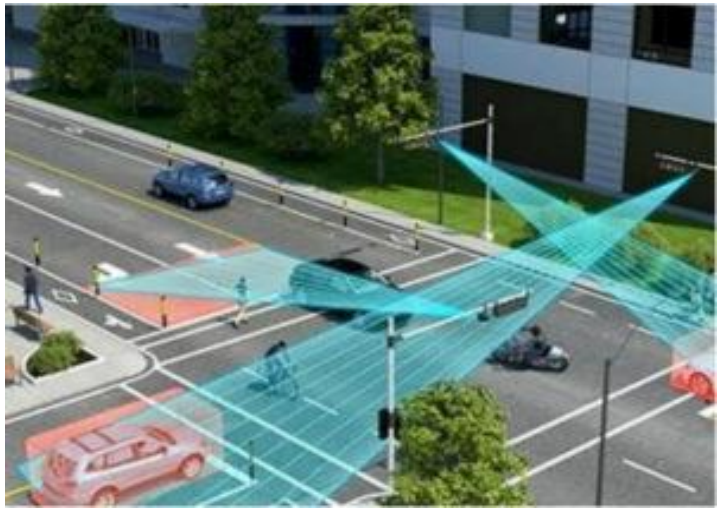
# Importance of Visibility Improvement for Video-based automated applications

**Dr. Prashant W. Patil**,
Assistant Professor, MFSDS&AI,
Indian Institute of Technology Guwahati, India

# Video based Automated Applications

- Automated video surveillance system.



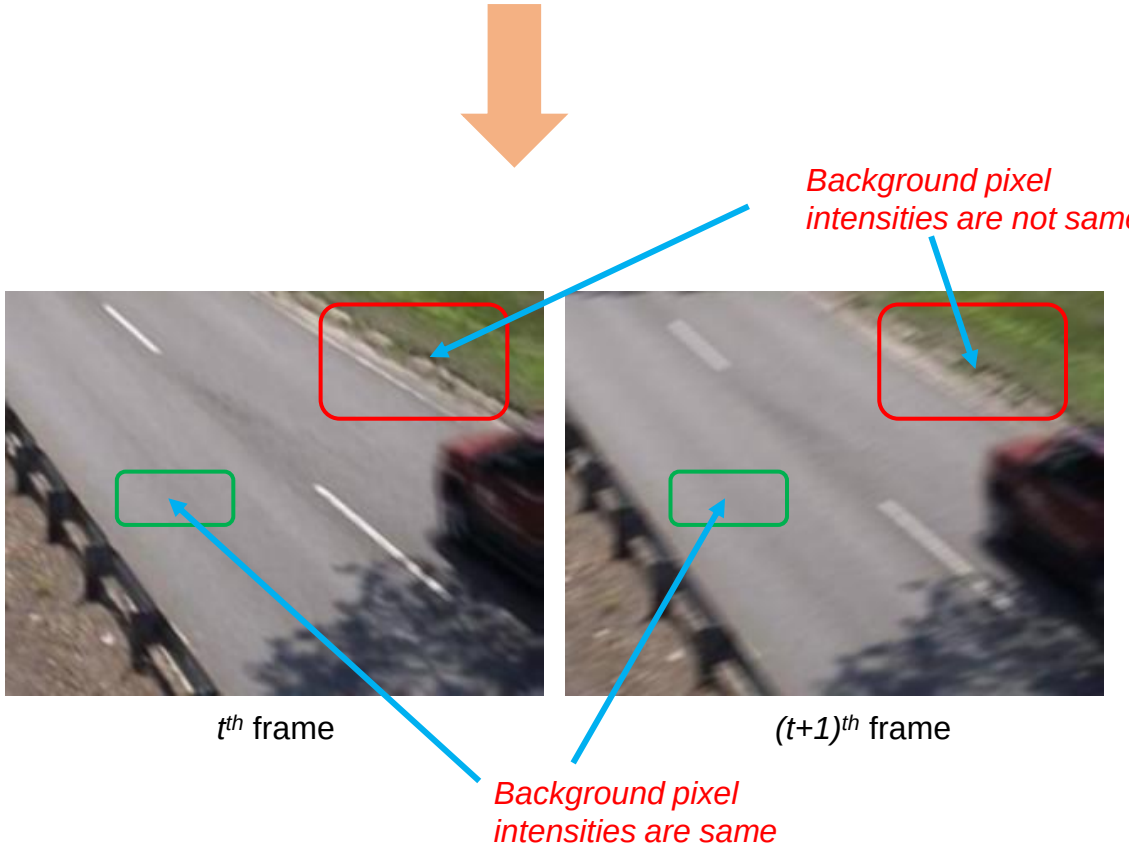
- Automated traffic monitoring.

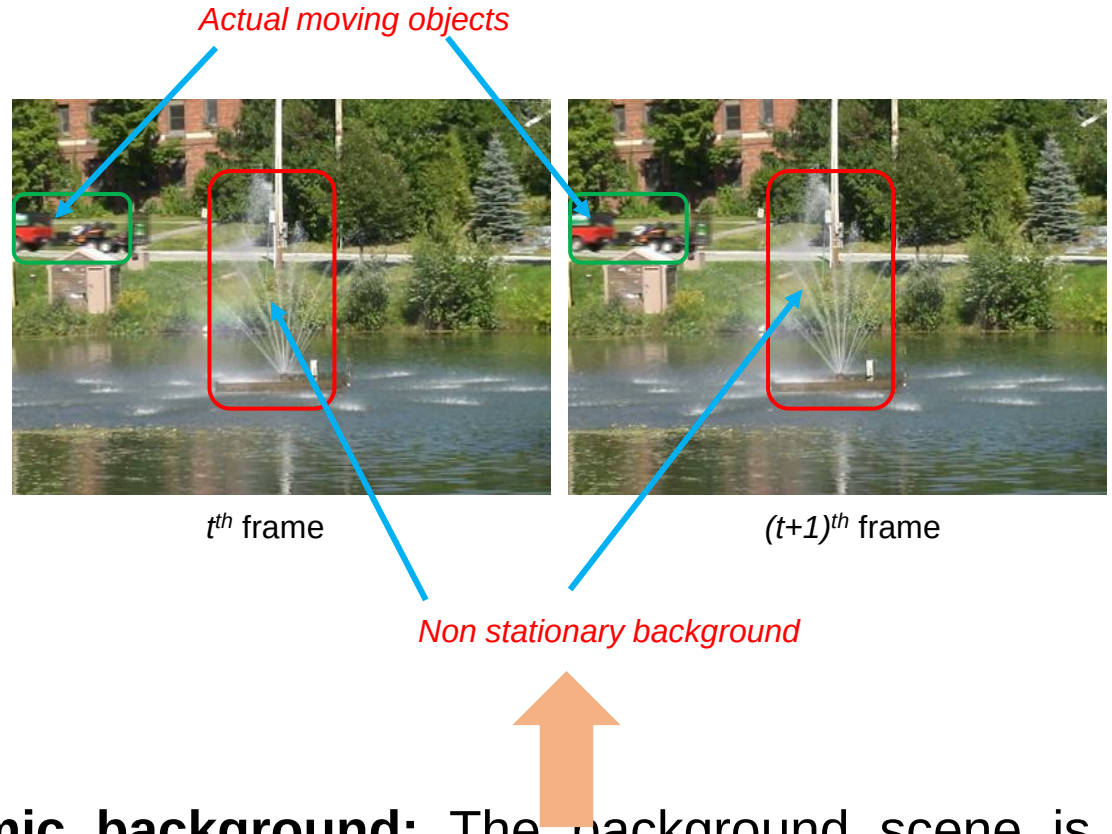
- Automated driving assistance system.

- **Expects clean data as input**
- **Heavily depends on sub-tasks**

# Major Challenges for Automated Surveillance

- **<u>Camera</u> <u>jitter</u>:** In some cases, camera may not be static, it may move frequently due to physical influence.

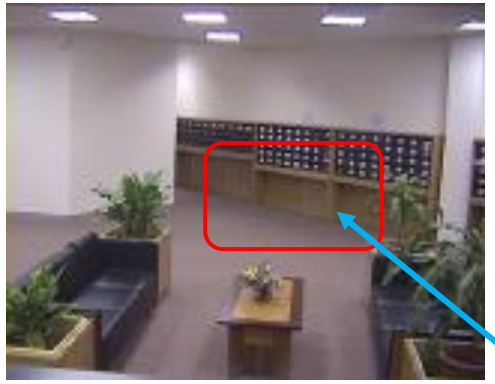


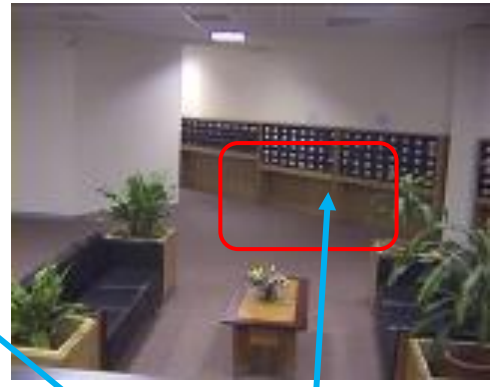

- **<u>Dynamic</u> <u>background</u>:** The background scene is rarely static due to movement in the background (*waving tree leaves*), especially in outdoor scenes.

# Major Challenges for Automated Surveillance

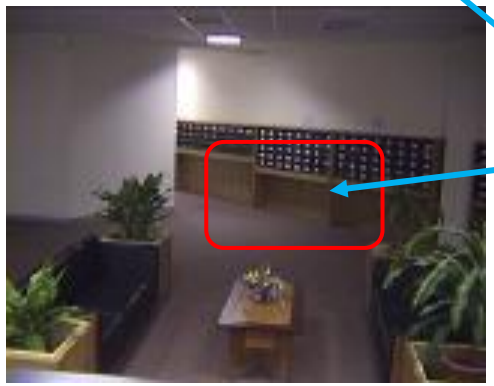
- **<u>Illumination changes</u>:** When scene lighting changes gradually (*e.g. moving clouds in the sky*) or instantly.



$t^{th}$ frame



$(t+10)^{th}$ frame



$(t+20)^{th}$ frame

*Background pixel intensities are changing with respect to time*

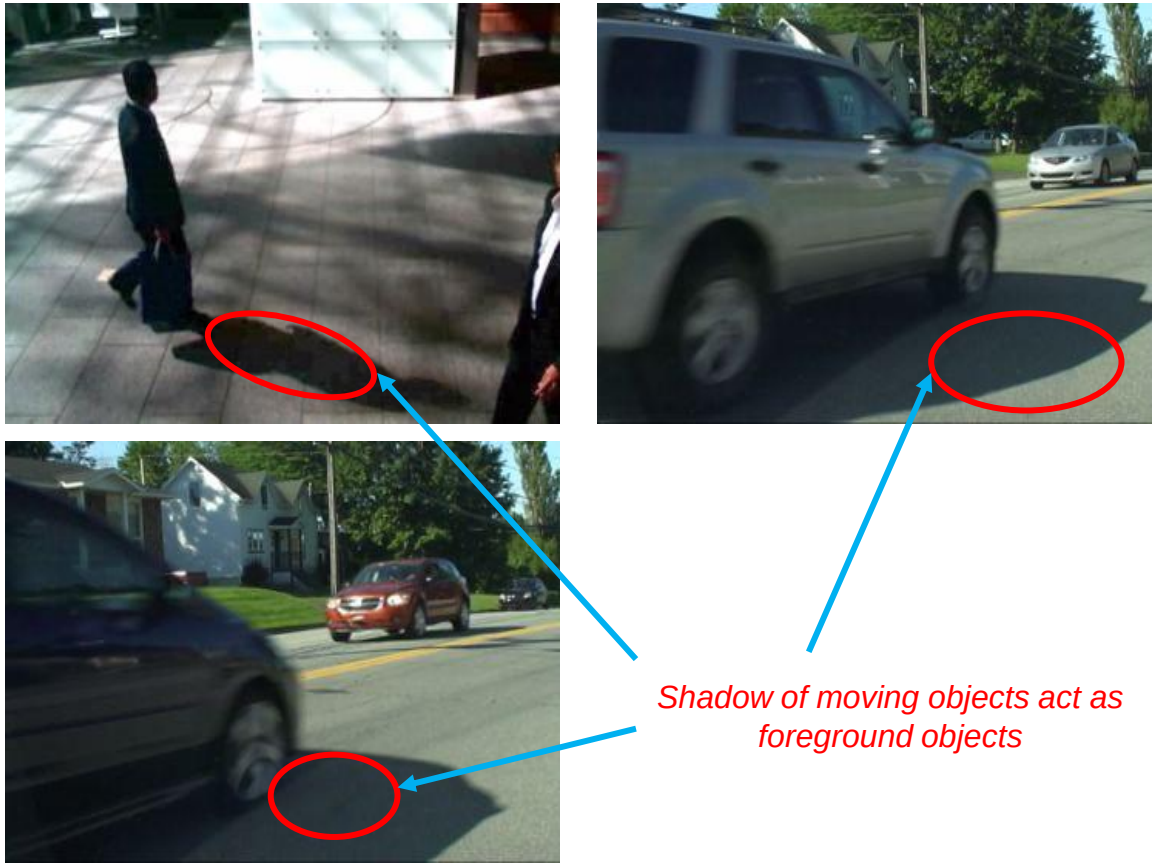
*Background and Foreground pixel intensities are same*

- **<u>Camouflage</u>**: Most of the background subtraction based algorithms work with pixel or color intensities.

# Major Challenges for Automated Surveillance

- **<u>Challenging</u> <u>Weather</u>**: Outdoor videos showing low-visibility winter storm conditions and the dark tire tracks left in the snow have potential to cause false positives

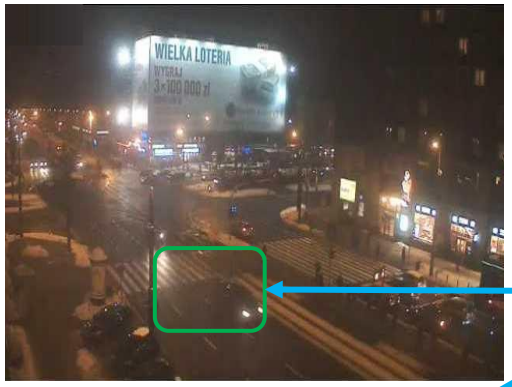
*Visibility of foreground objects and pixel intensities are not clear*



*Shadow of moving objects act as foreground objects*

- **<u>Shadows</u>**: Dark, moving shadows that do not fall under the illumination change category should not be labeled as foreground.

# Major Challenges for Automated Surveillance




*Foreground pixel / objects' intensities are not clear*

➢ **<u>Night</u> <u>Videos</u>**: Most of the pixels have a similar color in a night scene, recognition of foreground objects and their contours is difficult.

➢ **<u>Intermittent</u> <u>Object</u> <u>Motion</u>**: Foreground objects that are embedded into the background scene and start moving after back- ground initialization are the so-called ghosts

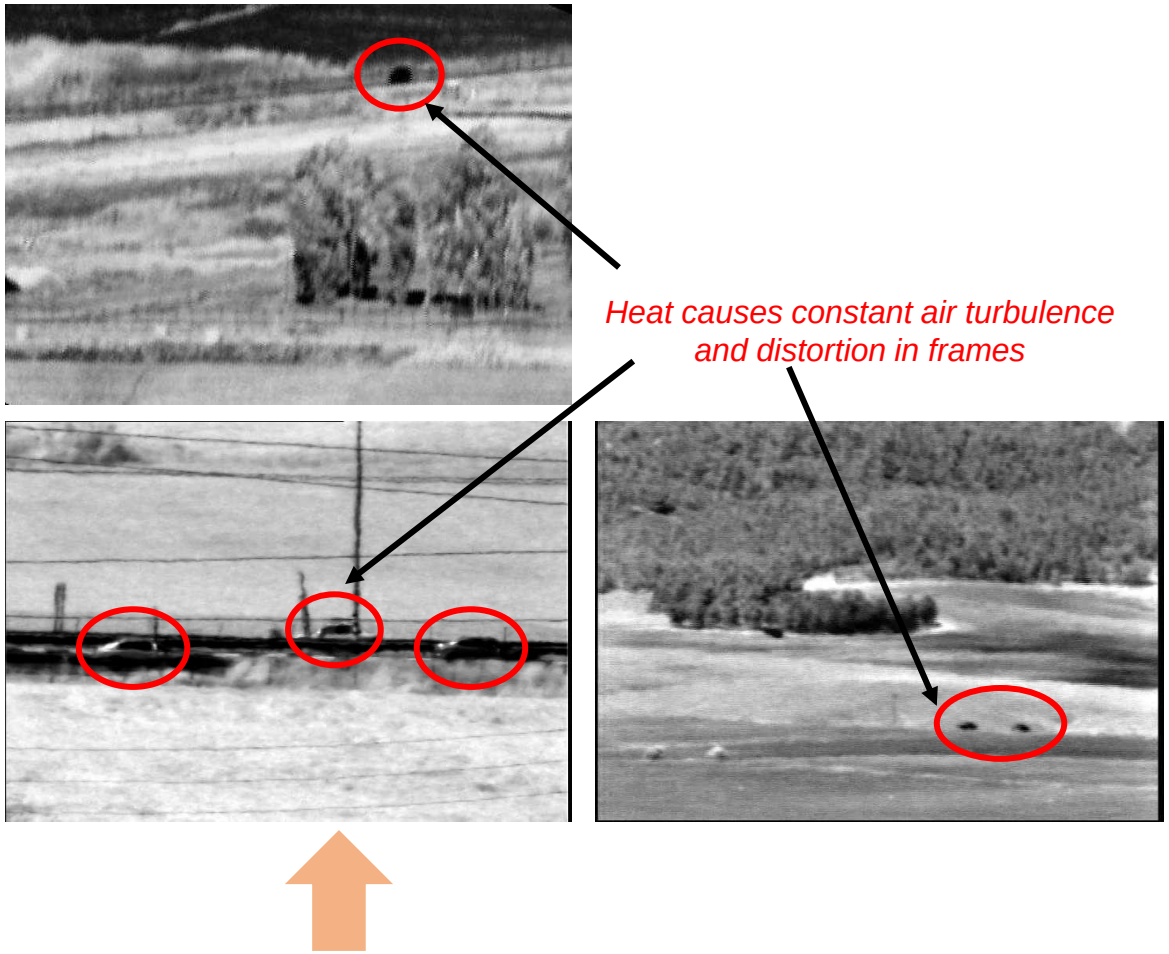

$t^{th}$ frame

*Foreground objects acts as background objects for particular span of time*


$(t+10)^{th}$ frame


$(t+20)^{th}$ frame

# Major Challenges for Automated Surveillance



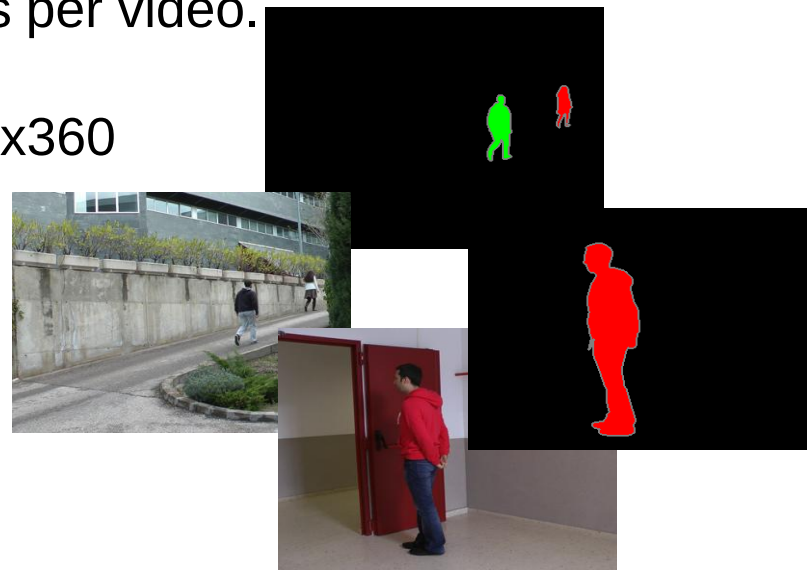

- **Turbulence**: The video recorded from long distance (5 to 15 km) and size of the objects are very small.

- **Bigger Moving Objects**: Foreground objects are having large size and share same texture feature.

# Experimental Databases

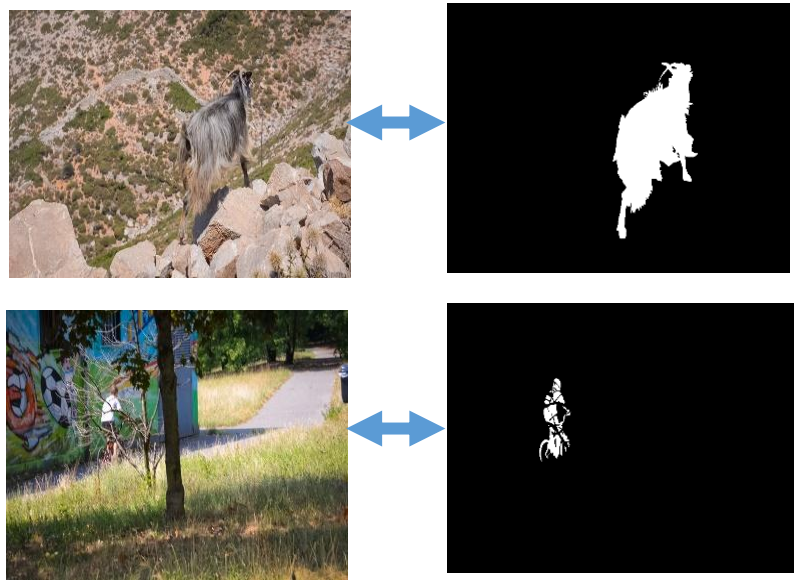
## LASIESTA Database [1]

- **Total 20 videos**
- Camouflage, rainy videos are covered.
- 100 to 400 frames per video.
- Size of frame 640x360



## DAVIS-2016 Database [2]

- **Total 50 videos**
- 50 to 200 frames per video.
- Size of frame 854x480



[1]. Cuevas *et al.* "*Labeled dataset for integral evaluation of moving object detection algorithms: LASIESTA*", CVIU-2016
[2]. Perazzi *et al.*, "A benchmark dataset and evaluation methodology for video object segmentation," CVPR, 2016.

# General Pipeline for Moving Object Segmentation

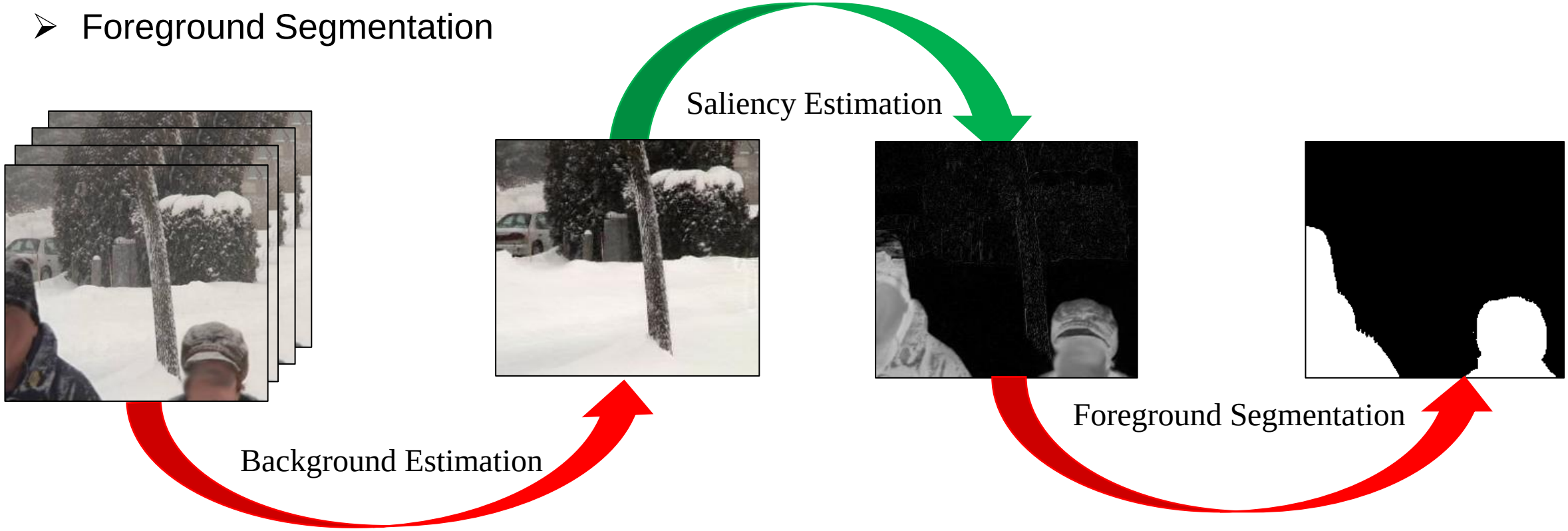
There are three main steps involved in any background-foreground separation algorithm:

- Background Estimation
- Motion Saliency Estimation
- Foreground Segmentation

# General Pipeline for Moving Object Segmentation

There are three main steps involved in any background-foreground separation algorithm:

- Background Estimation
- Motion Saliency Estimation
- Foreground Segmentation



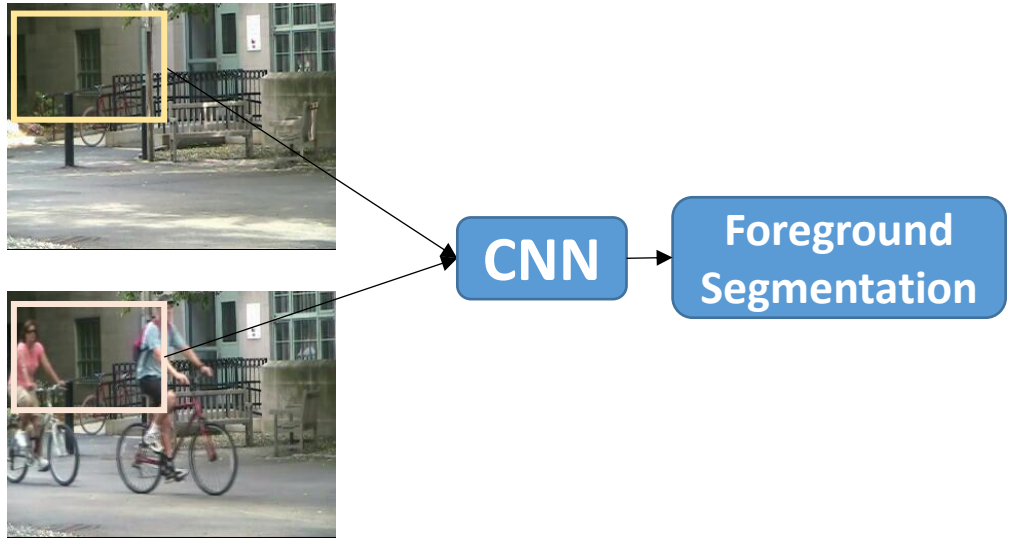

Generalized architecture of MOS approach.

# Discussion on some existing methods

- Background subtraction based methods

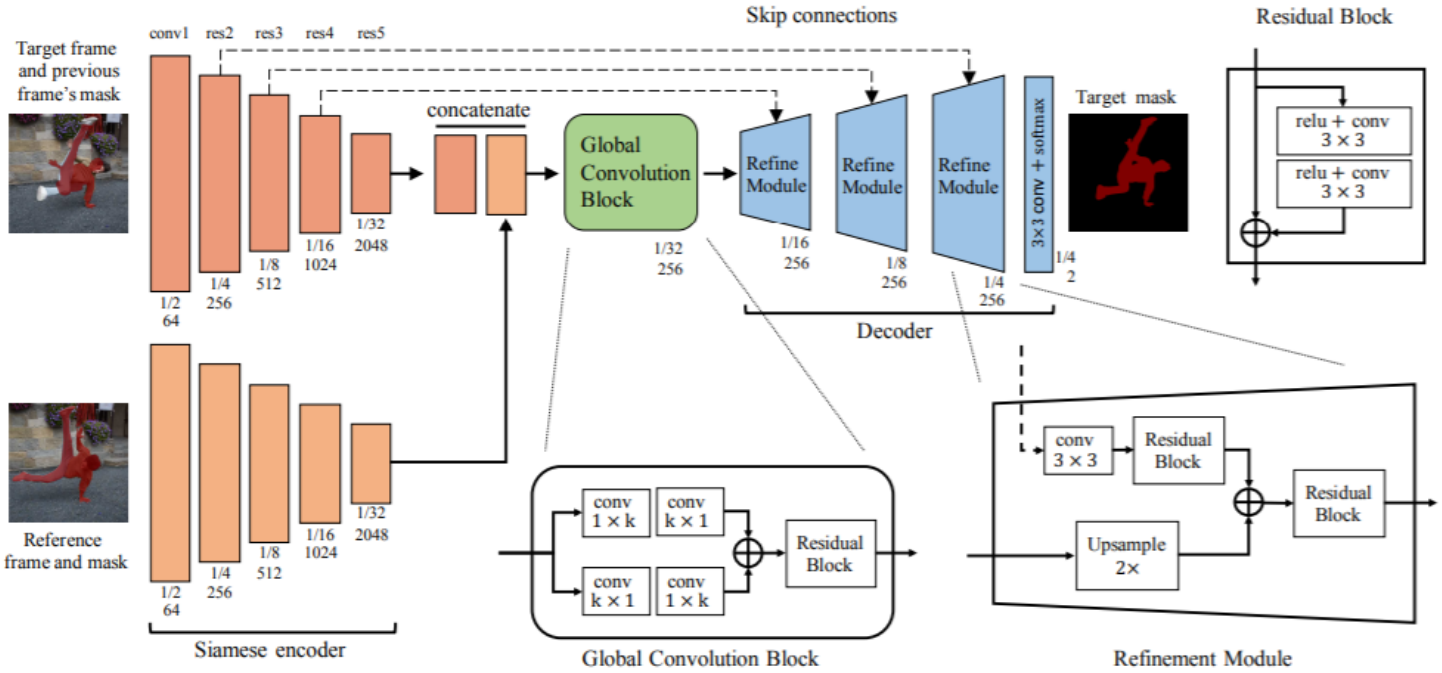
# Discussion on some existing methods

- Background subtraction based methods
- Background learning based methods

# Discussion on some existing methods

- Background subtraction based methods
- Background learning based methods
- Reference frame based methods



# Discussion on some existing methods

- Background subtraction based methods
- Background learning based methods
- Reference frame based methods
- Scene dependent based methods
- Frame-to-frame learning based methods

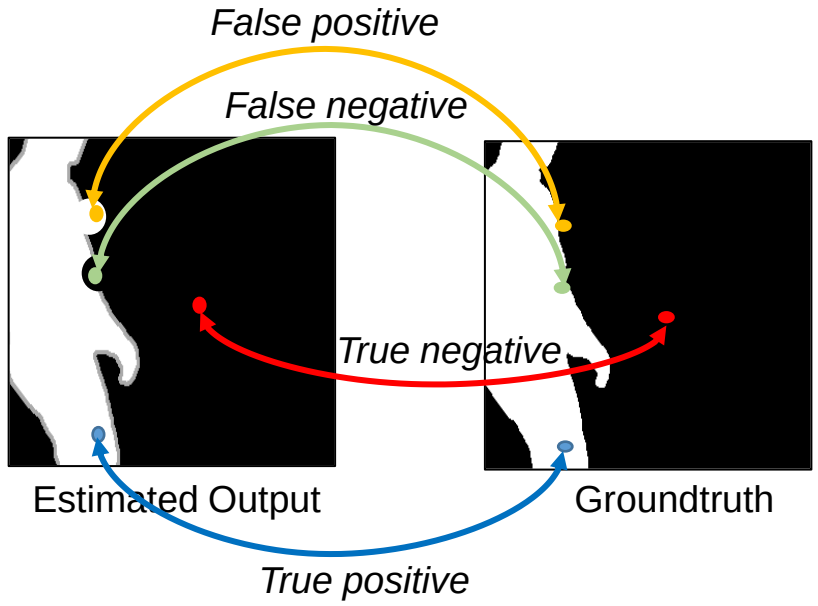
# Evaluation Measures

- The average {Precision, Recall and F-measure} are the evaluation measures for MOS.
- These parameters are  measured in terms of True positive, True negative, False positive, False negative.

$$\text{Precision} = \frac{True\ Positive}{True\ Positive + False\ Positive}$$

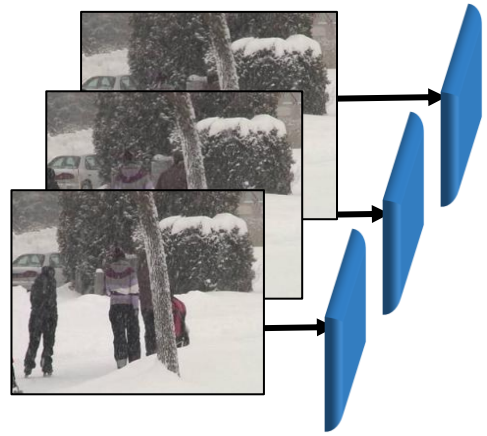
$$\text{Recall} = \frac{True\ Positive}{True\ Positive + False\ Negative}$$

$$F - measure = \frac{2 \times Precision \times Recall}{Precision + Recall}$$



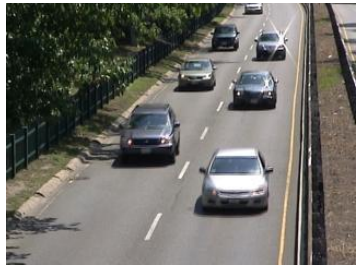
# Observations

1) Instead of processing consecutive frames through single encoder, we have designed the parallel processing of consecutive frames through different encoder network along with recurrent feature sharing.



Parallel processing of input frames

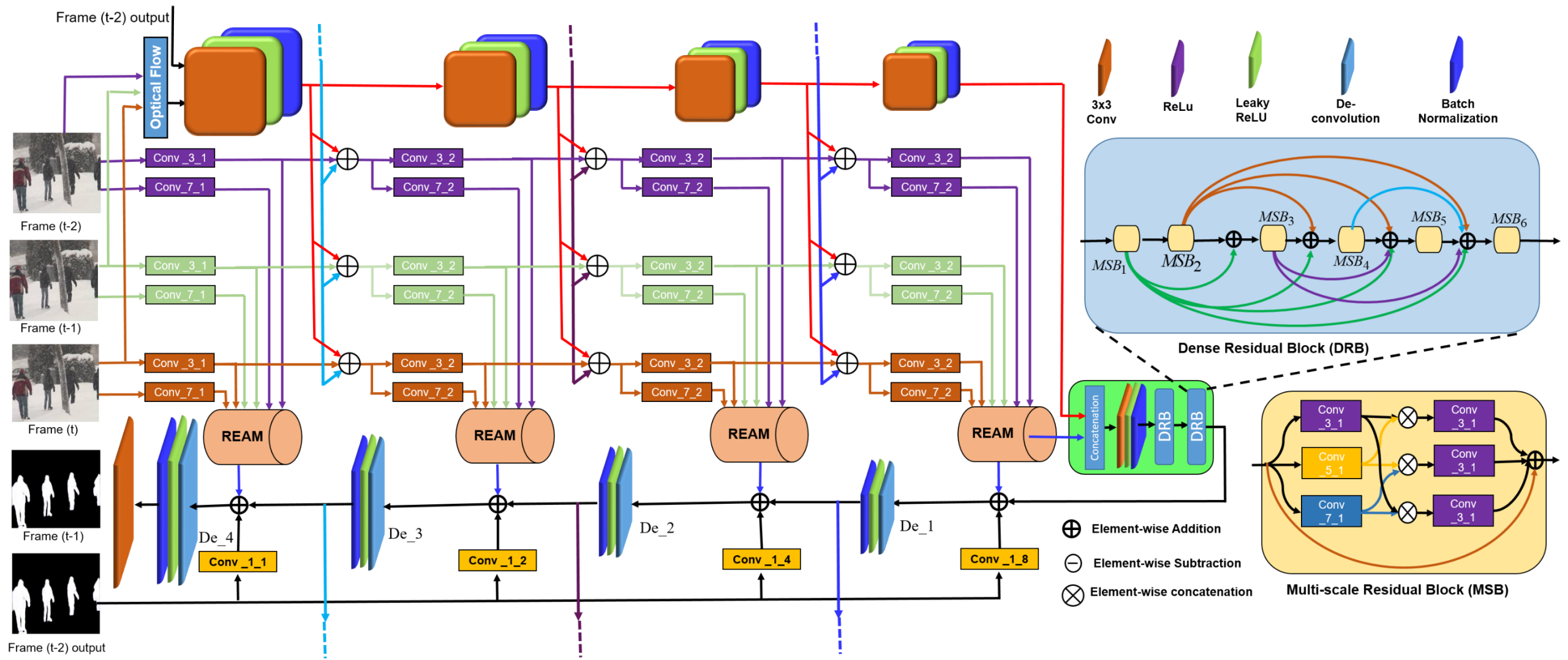
2) Motion between two consecutive frames is very small, the feedback of previous frame learned decoder features may be effective for consistent foreground segmentation.



$(t-1)^{th}$ frame

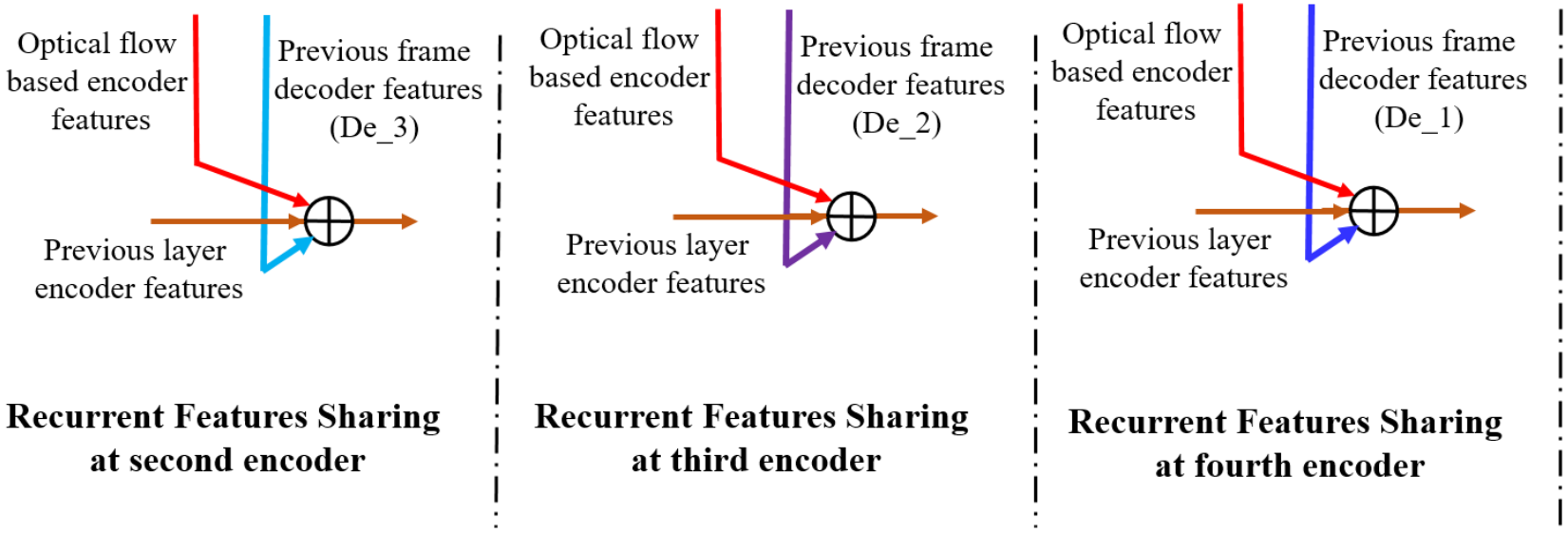
$t^{th}$ frame

# Proposed Solution

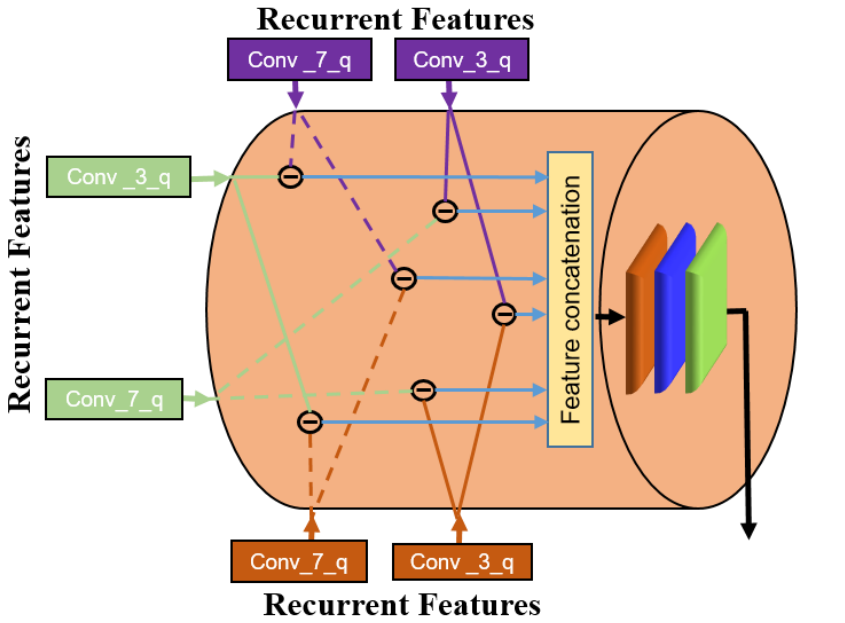



1. **Patil *et. al.*,** "An unified recurrent video object segmentation framework for various surveillance environments", IEEE TIP-2021 (IF-11.041).

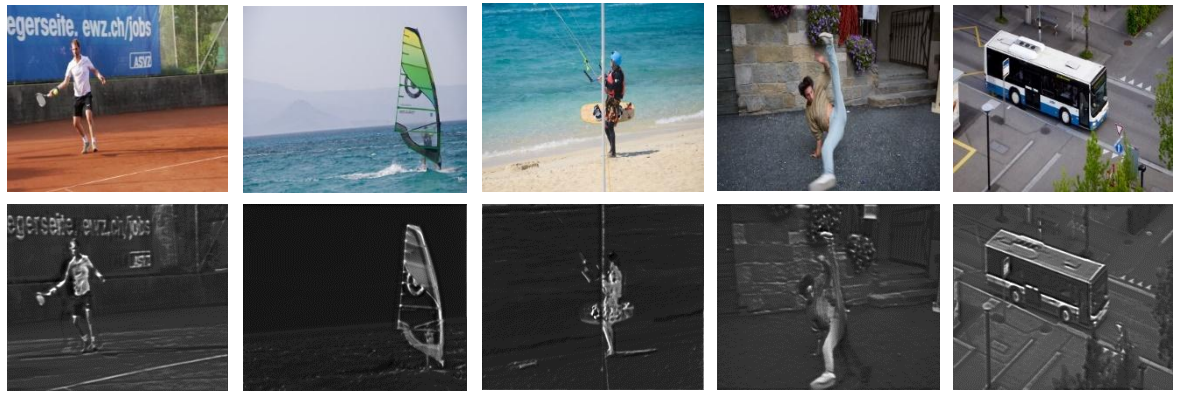
# Proposed Solution




(a) Proposed recurrent Features Sharing at each encoder




(b) Proposed Recurrent Edge Aggregation Module



# Training Details

## Global Training-Testing

- Video frames are divided into training-testing sets.
- CDnet-2014 database is used.

Train

Test

## Disjoint Training-Testing

- Videos are divided into training-testing sets
- DAVIS-2016 and SegTrack-V2 databases are used.

Train

Test

## Cross-Database Training-Testing

- Databases are divided into training-testing
- LASIESTA, GTFD and CDnet-2014 (Thermal Category) databases are used.

# Result Analysis

## Quantitative Result Analysis

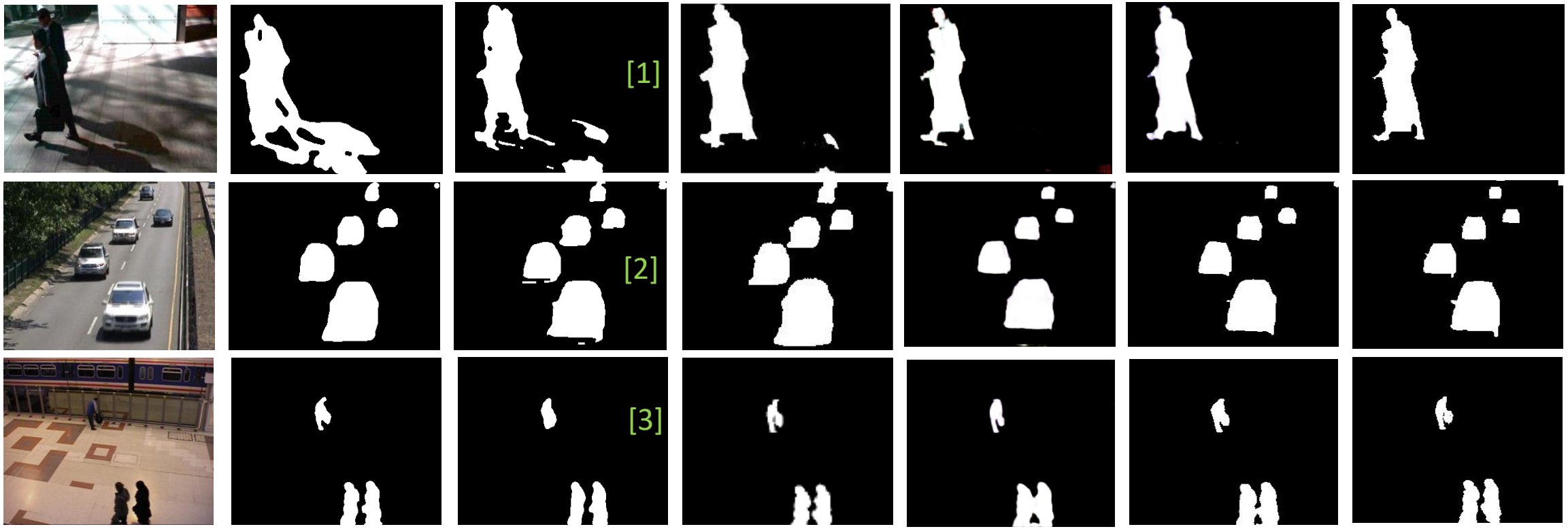
Quantitative result analysis on all the databases

| Database | Training-Testing | Existing Method ($2^{nd}$ Best) | Proposed Method-I | Proposed Method-II |
|---|---|---|---|---|
| DAVIS-16 | Disjoint | 0.902 [1] | **0.915** | **0.938** |
| SegTrack-v2 | Disjoint | 0.899 [2] | **0.918** | **0.929** |
| CDnet-14 | Local | 0.964 [3] | **0.969** | **0.973** |
| LASIESTA | Cross | 0.906 [4] | **0.910** | **0.940** |
| AGVS | Transfer | 0.710 [5] | **0.740** | **0.790** |
| GTFD | Cross | 0.730 [6] | **0.750** | **0.780** |


[1]. Li *et al*., “Motion guided attention for video salient object detection,” CVPR 2019.
[2]. Zhuo *et al*., “Unsupervised online video object segmentation with motion property understanding”, IEEE TIP 2019
[3]. Akilan *et al*., “A 3D CNN-LSTM based image-to-image foreground segmentation”. IEEE ITS 2019
[4]. Zhao *et al*., “Dynamic deep pixel distribution learning for background subtraction”, IEEE CSVT 2019
[5]. Zhang *et al*., “Extended motion diffusion-based change detection for airport ground surveillance”, IEEE TIP 2020
[6]. Li *et al*., “Weighted low rank decomposition for robust grayscale-thermal foreground detection”, IEEE CSVT 2017

# Result Analysis

## Global Training-Testing Result Analysis

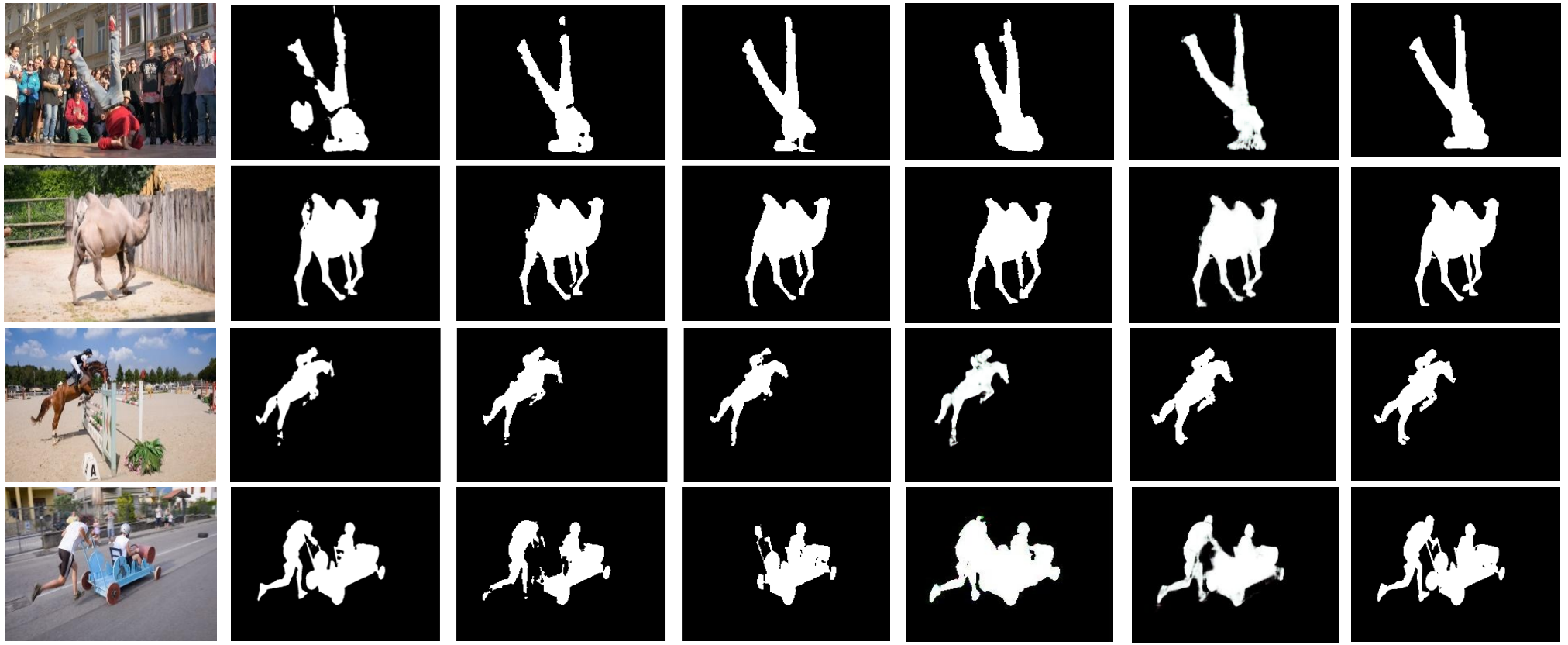



Visual results comparison on CDnet-2014 database


[1]. Akilan *et al*., “Video foreground extraction using multi-view receptive field and encoder–decoder DCNN for traffic and surveillance applications,” IEEE TVT 2019.
[2]. Akilan *et al*., “sendec: An improved image to image cnn for foreground localization”, IEEE TITS 2019
[3]. Akilan *et al*., “3d cnn-lstmbased image-to-image foreground segmentation”. IEEE TITS 2019
[4]. Babee *et al*., “A deep convolutional neural network for video sequence background subtraction,” PR 2018
[5]. Patil *et al*., “Msfgnet: A novel compact end-to-end deep network for moving object detection”, IEEE TITS 2018

# Result Analysis

## Disjoint Training-Testing Result Analysis

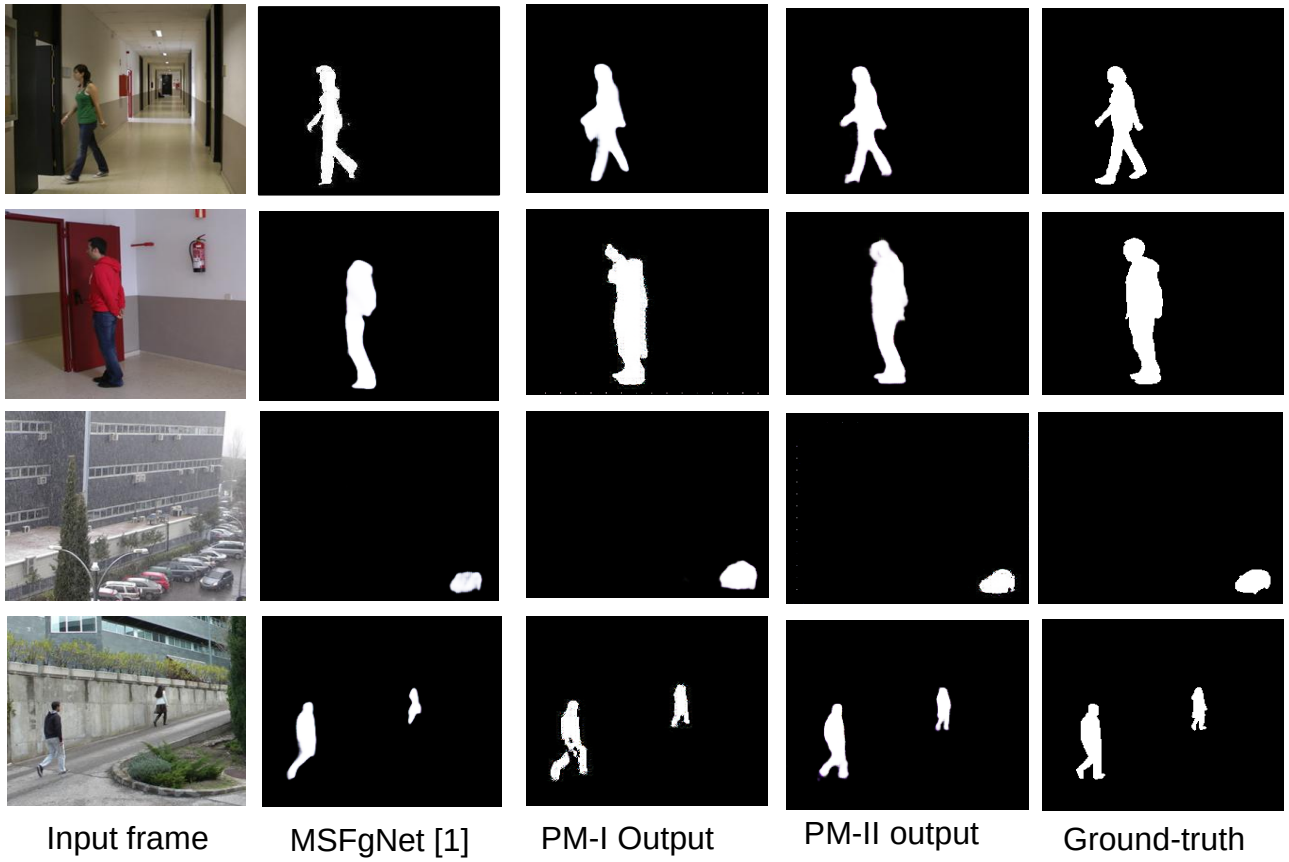



Visual results comparison on DAVIS-2016 database

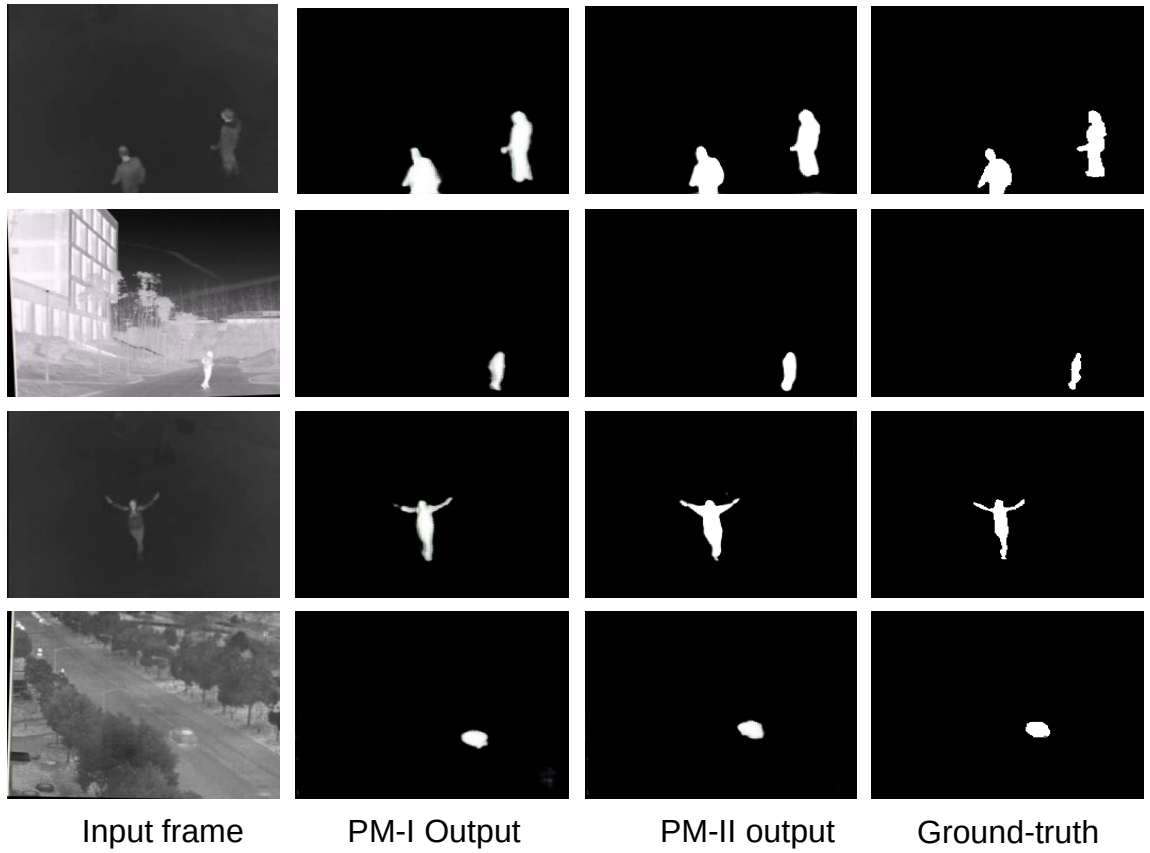

[1]. Wang *et al*., “Ranet: Ranking attention network for fast video object segmentation,” CVPR 2019.
[2]. Maninis *et al*., “Video object segmentation without temporal information”, IEEE TPAMI 2019
[3]. Juiten *et al*., “Premvos: Proposal-generation refinement and merging for video object segmentation”. ACCV 2018

# Result Analysis

## Cross Training-Testing Result Analysis




Visual results comparison on LASIESTA database

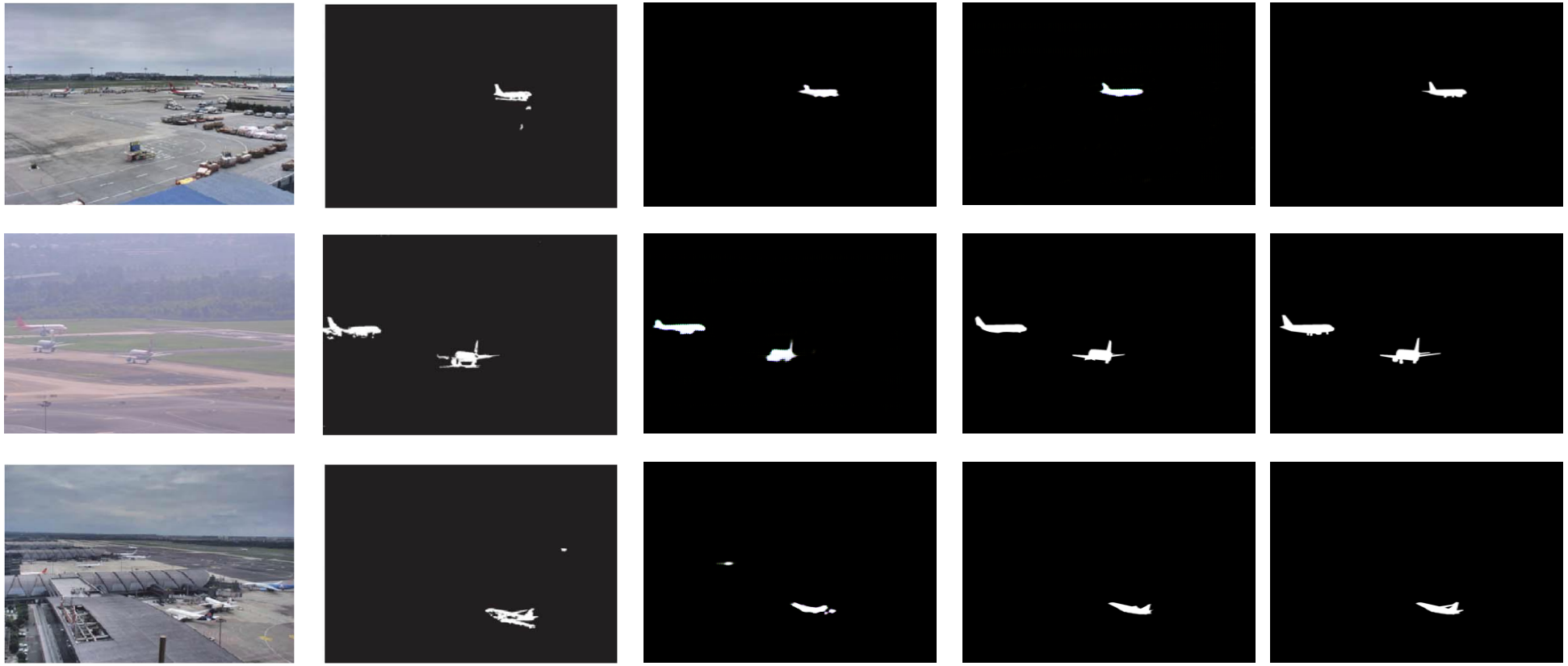



Visual results comparison on GTFD database

[1]. Patil *et al*., “Msfgnet: A novel compact end-to-end deep network for moving object detection”, IEEE TITS 2018

# Result Analysis

## Fine Tuning Based Result Analysis :



Input Frame | [1] | PM-I | PM-II | Ground-truth

Visual results comparison on AGVS database

[1]. Zhang *et al*., “Extended motion diffusion-based change detection for airport ground surveillance”, IEEE TIP 2020

# Result Analysis

## Ablation Study :

- Main Blocks of the proposed architecture: Encoder Block, Recurrent Edge Aggregation Module (REAM), Motion refinement Block (MRB), Decoder with and without Feedback and Fusion of different features

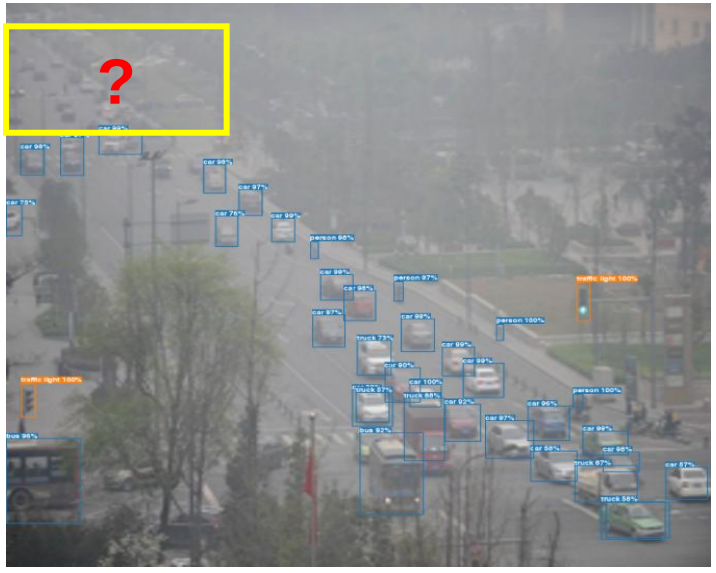
Different parameter analysis of the proposed network (*W/i: with, W/o: without, FB: feedback, MFeat: optical flow based features, RFeat: previous frame decoder features shared recurrently to current frame encoder features*)

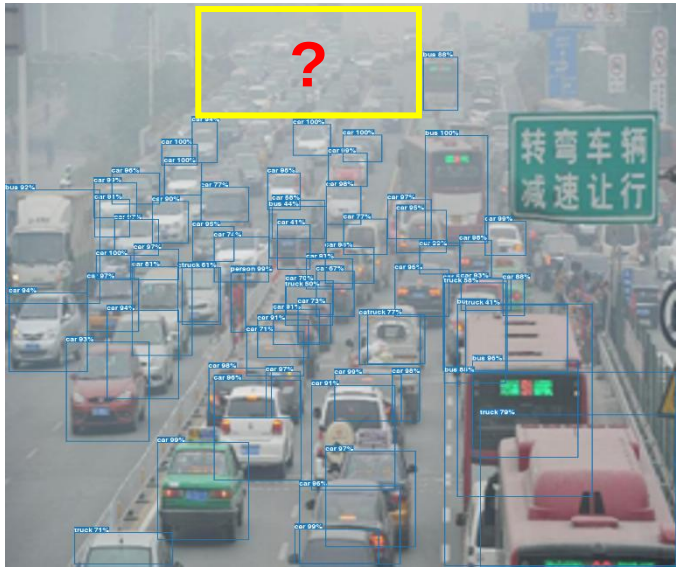
| Networks | Encoder | | Decoder | | REAM | MRB | | Fusion | | F-measure | MAE |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | RGB Based | OF based | W/i FB | W/o FB | | W/i | W/o | MFeat | RFeat | | |
| Network I | ✓ | - | ✓ | - | - | ✓ | - | - | - | 0.7885 | 0.0489 |
| Network II | ✓ | - | ✓ | - | ✓ | ✓ | - | - | - | 0.8246 | 0.0395 |
| Network III | - | ✓ | ✓ | - | - | ✓ | - | - | - | 0.8648 | 0.0296 |
| Network IV | ✓ | ✓ | - | ✓ | - | - | ✓ | - | - | 0.8691 | 0.0274 |
| Network V | ✓ | ✓ | ✓ | - | - | ✓ | - | - | - | 0.8792 | 0.0216 |
| Network VI | ✓ | ✓ | ✓ | - | ✓ | ✓ | - | - | - | 0.9149 | 0.0191 |
| Network VII | ✓ | ✓ | ✓ | - | ✓ | ✓ | - | - | ✓ | 0.9234 | 0.0182 |
| Network VIII | ✓ | ✓ | ✓ | - | ✓ | ✓ | - | ✓ | - | 0.9295 | 0.017 |
| Final Network | ✓ | ✓ | ✓ | - | ✓ | ✓ | - | ✓ | ✓ | **0.9387** | **0.0161** |

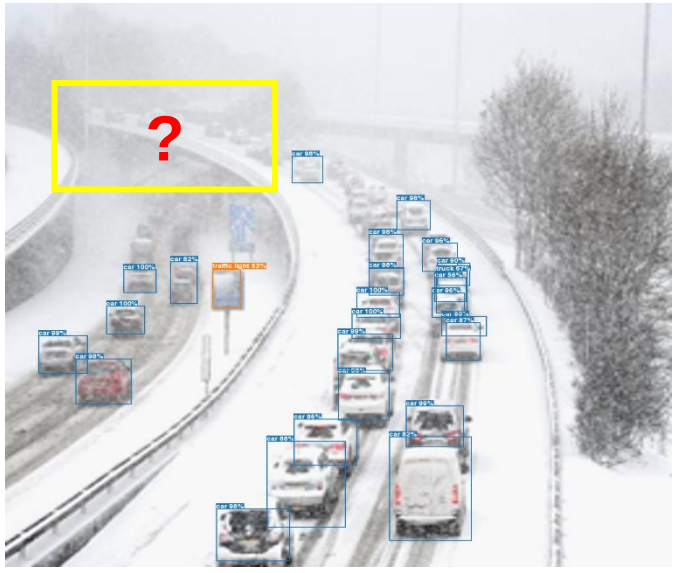
# Conclusion

- It is observed that the proposed inherent recurrent correlation learning-based EEM and DRBs are working effectively for moving object segmentation.
- Experimental analysis is carried out with different training-testing like global, disjoint, cross and transfer learning based configurations.
- From result analysis, it is evident that the proposed network outperforms the existing methods.

# Limitation of Object Detection/Segmentation

# Limitation of Object Detection/Segmentation

# Incident due to bad weather

# Incident due to bad weather

# Various Real World Weather Degradations

## Hazy Weather Conditions




# Various Real World Weather Degradations

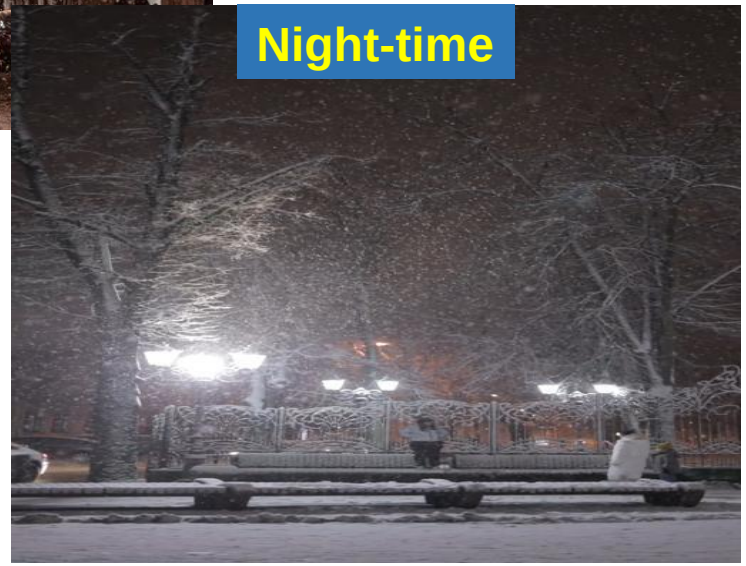
## Rainy Weather Conditions




# Various Real World Weather Degradations

## Snowy Weather Conditions

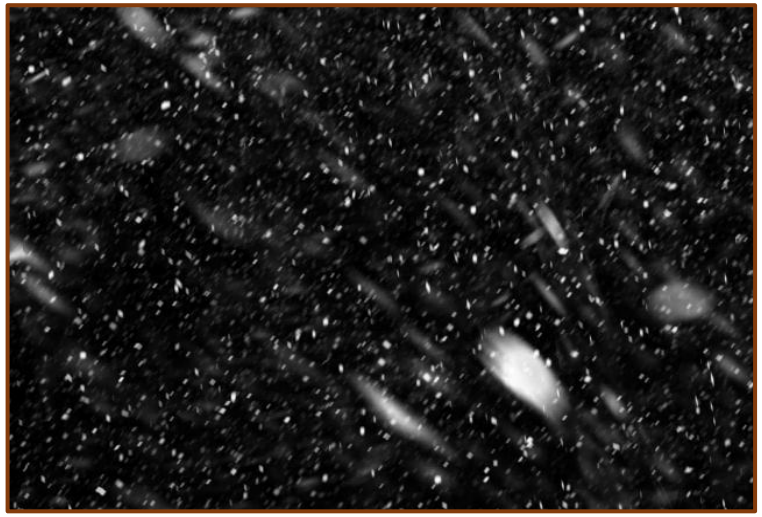


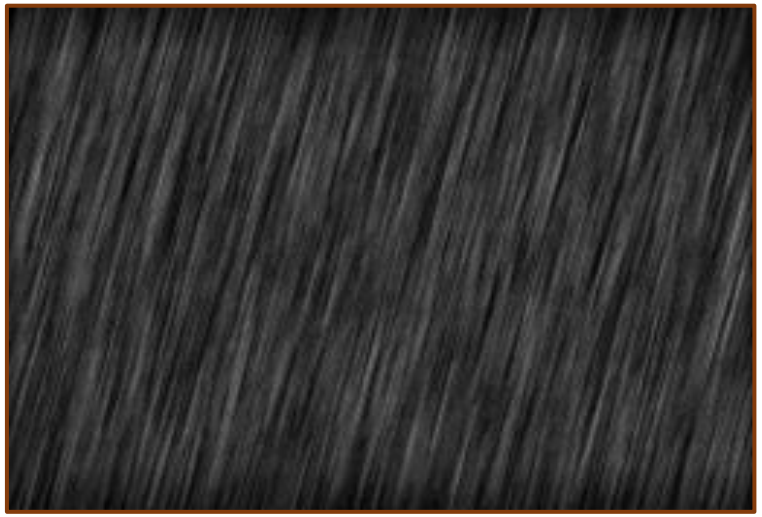



# Discussion on some existing methods

- Prior based models.





# Discussion on some existing methods

- Prior based models.

- Weather specific models

# Discussion on some existing methods

- Prior based models.

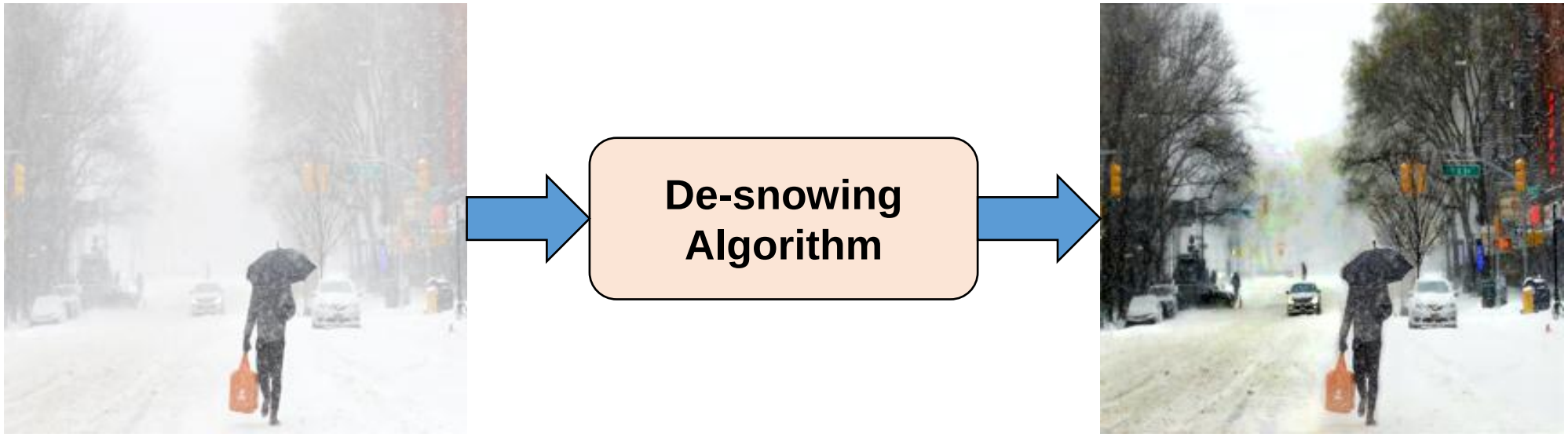
- Weather specific models

# Discussion on some existing methods

- Prior based models.
- Weather specific models.
- Sequential degradation learning
- Computational complexity.
- Limited video based algorithms.

# Brief Overview

- A custom-designed architecture for each weather condition with and without domain-specific knowledge.

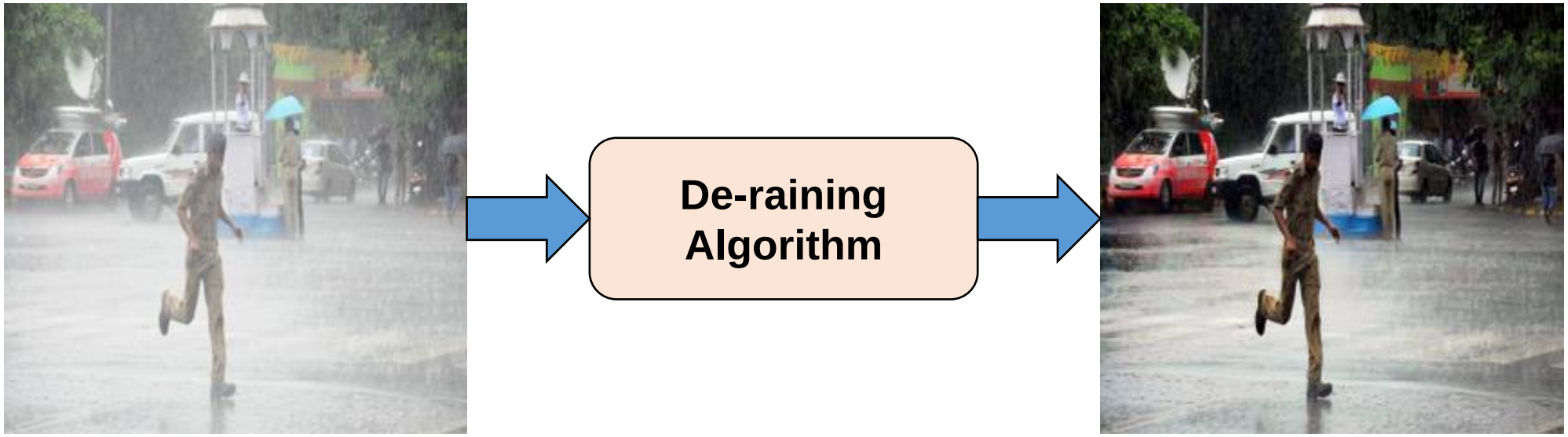



Input Image

Output Image

# Brief Overview

- A custom-designed architecture for each weather condition with and without domain-specific knowledge.

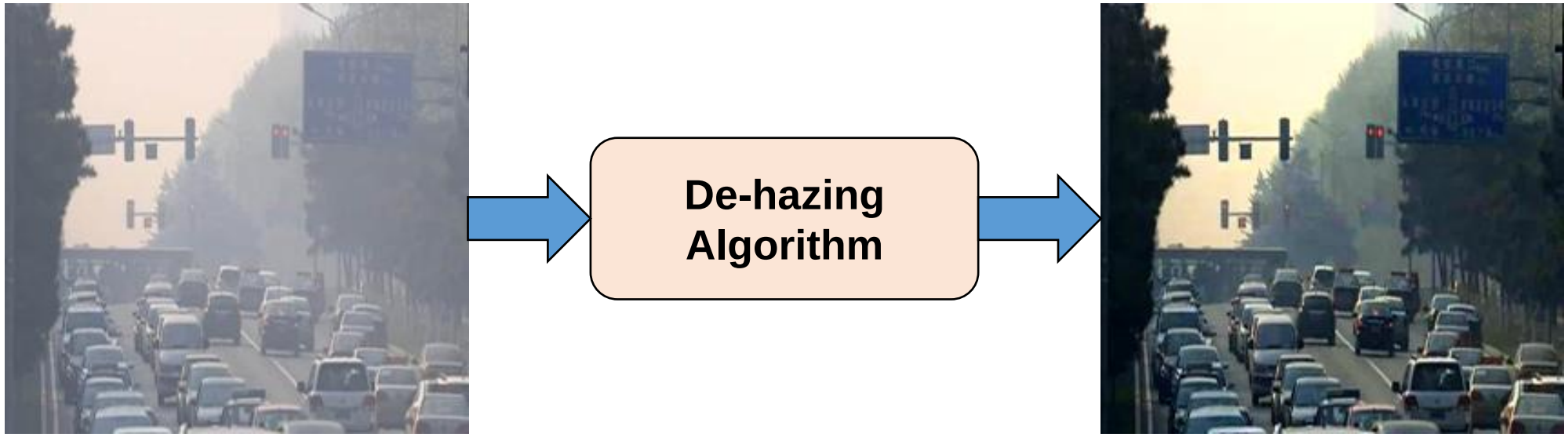



Input Image

Output Image

# Brief Overview

- A custom-designed architecture for each weather condition with and without domain-specific knowledge.



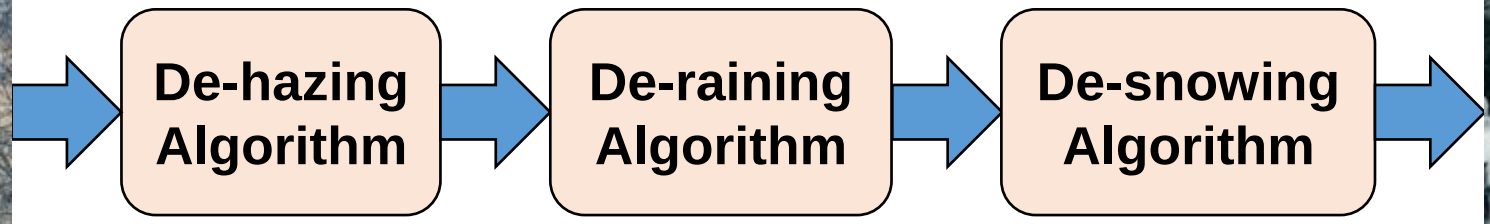

Input Image

Output Image

# Brief Overview

- A custom-designed architecture for each weather condition with and without domain-specific knowledge.

Input Image




Output Image

**Computational Complexity ?**

**Training data requirement ??**

**Sequential degradation learning ???**

# Why synthetic data generation?

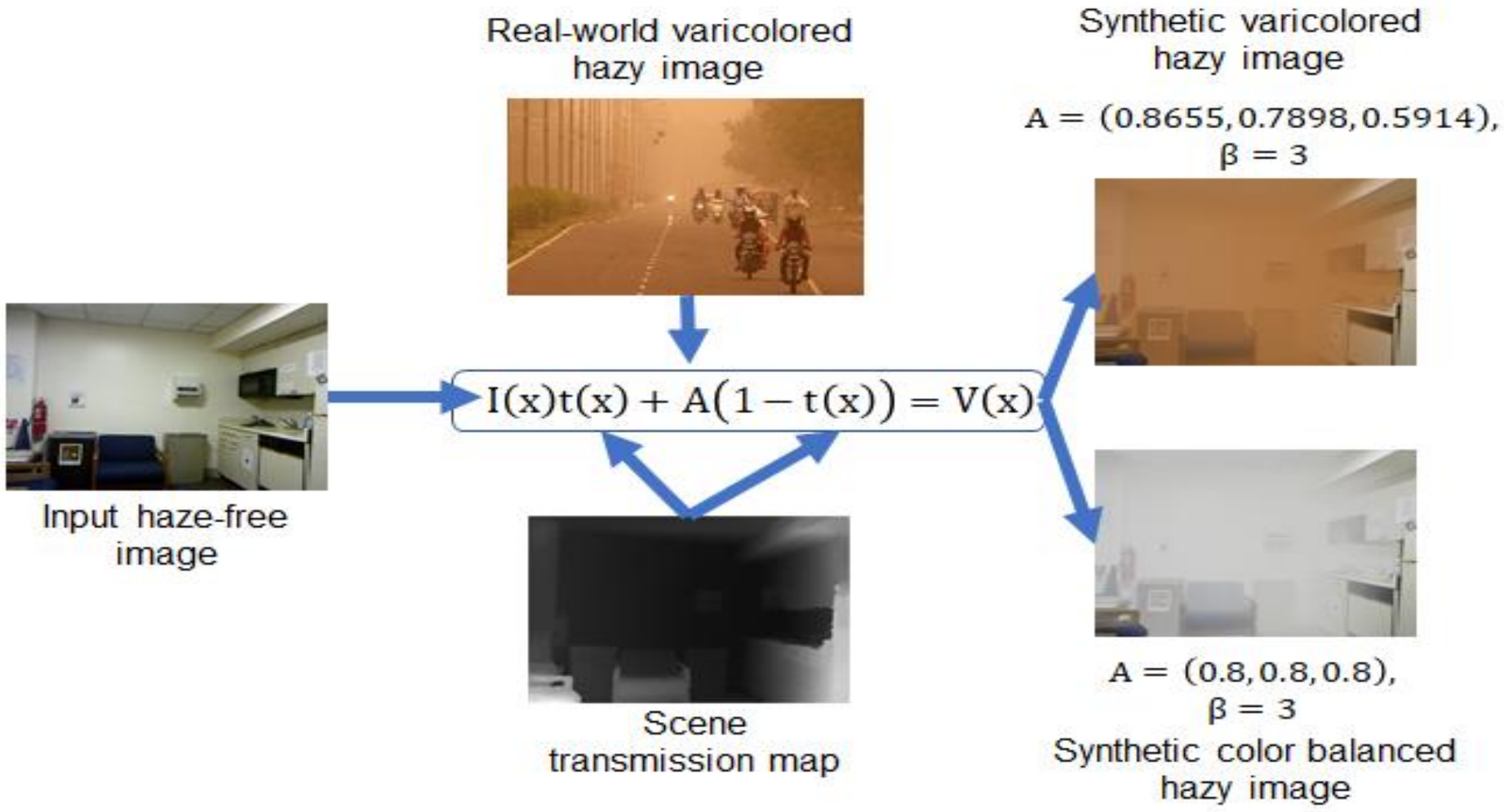
Input Images

Clean Images

? ? ?

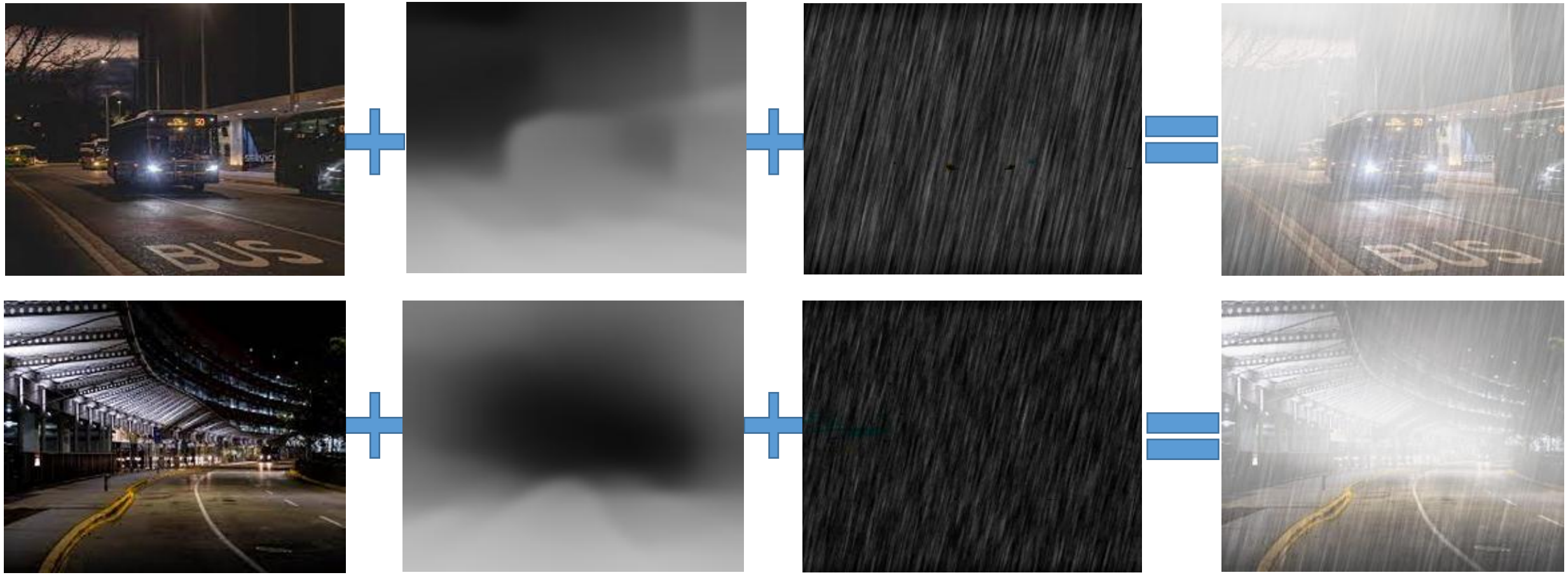
# Synthetic Image Generation

## Synthetic Hazy Image Generation



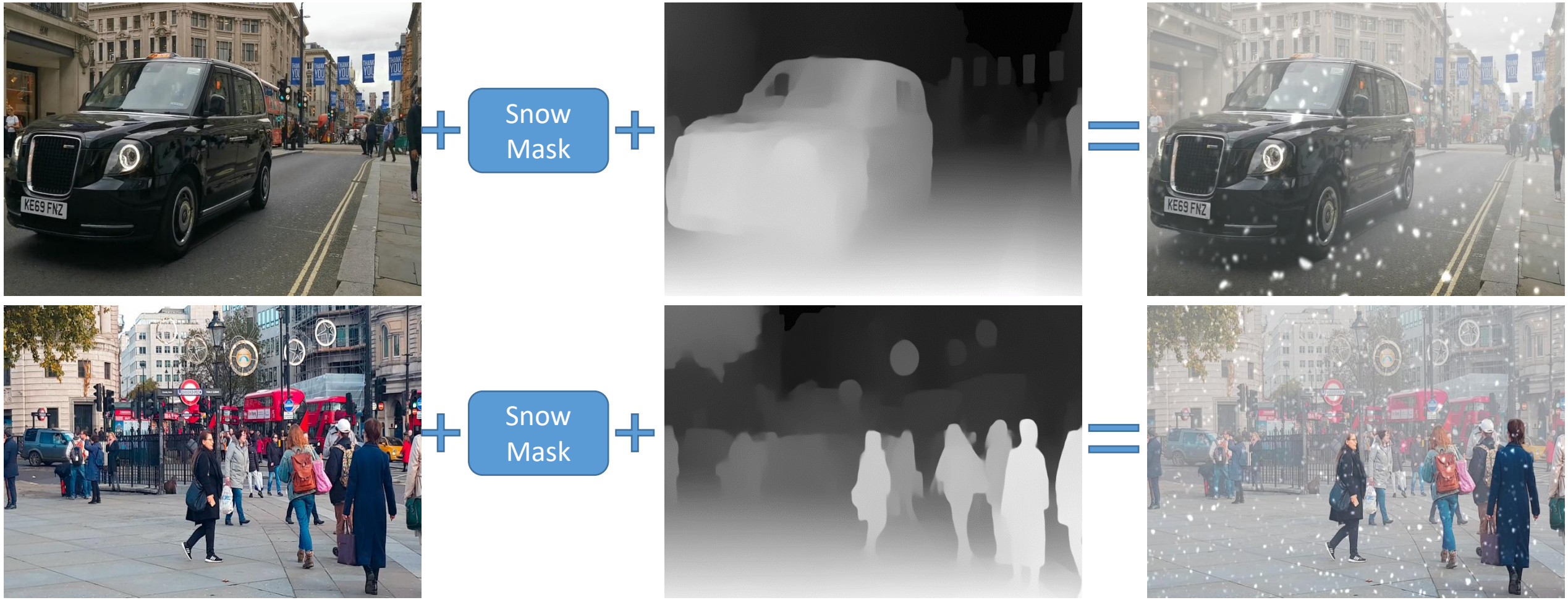
# Synthetic Image Generation

## Synthetic Rainy Image Generation



# Synthetic Image Generation

## Synthetic Snowy Image Generation




Input Image

Transmission Map

Snowy Image

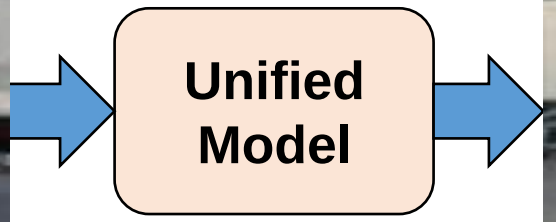
# Synthetic and Real data

Synthetic Images

Real-world Images

# Issue of Sequential degradation learning

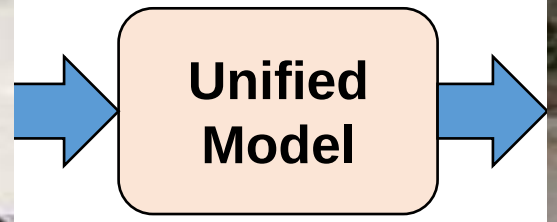
Input Image




Restored Image

# Issue of Sequential degradation learning

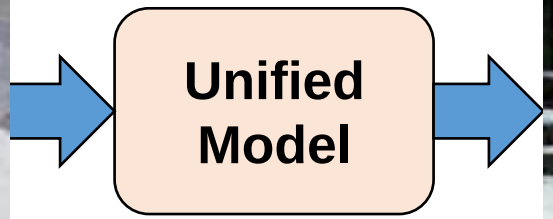
Input Image




Restored Image

# Issue of Sequential degradation learning

Input Image




Restored Image

# Complicated Weather Situations

# Complicated Weather Situations

Hazy Image

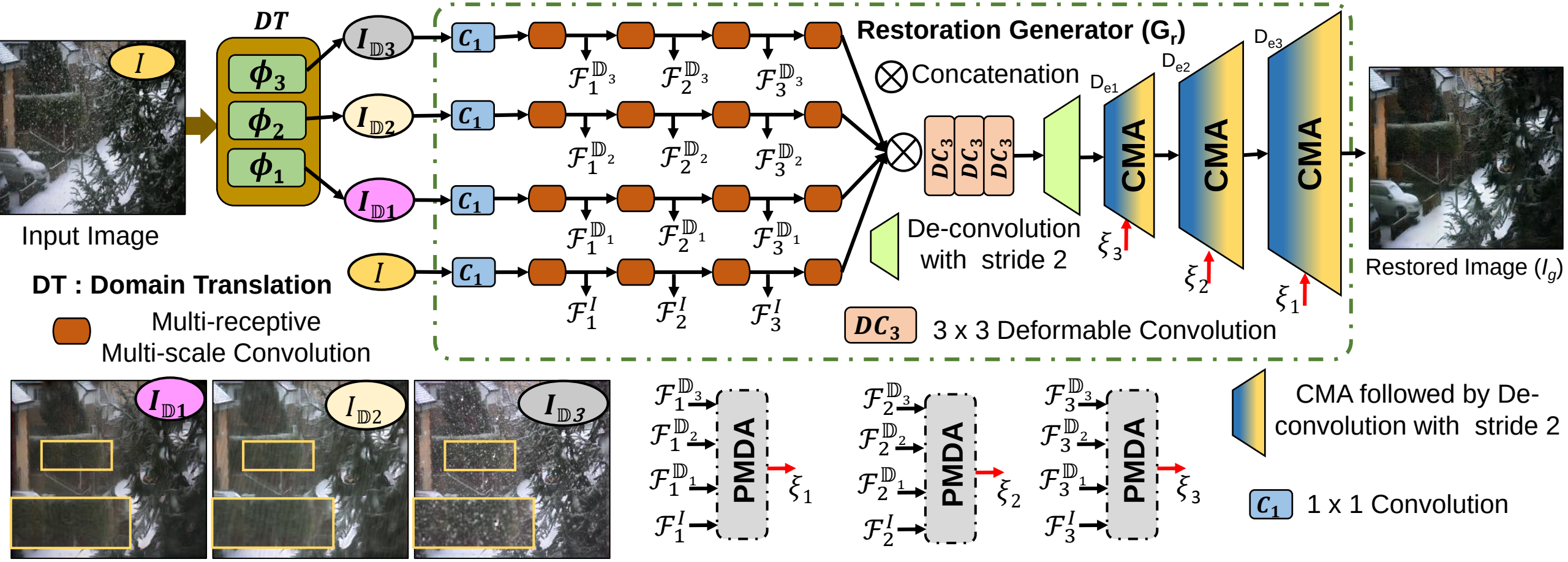
Rainy Image

Snowy Image

# Major Contributions

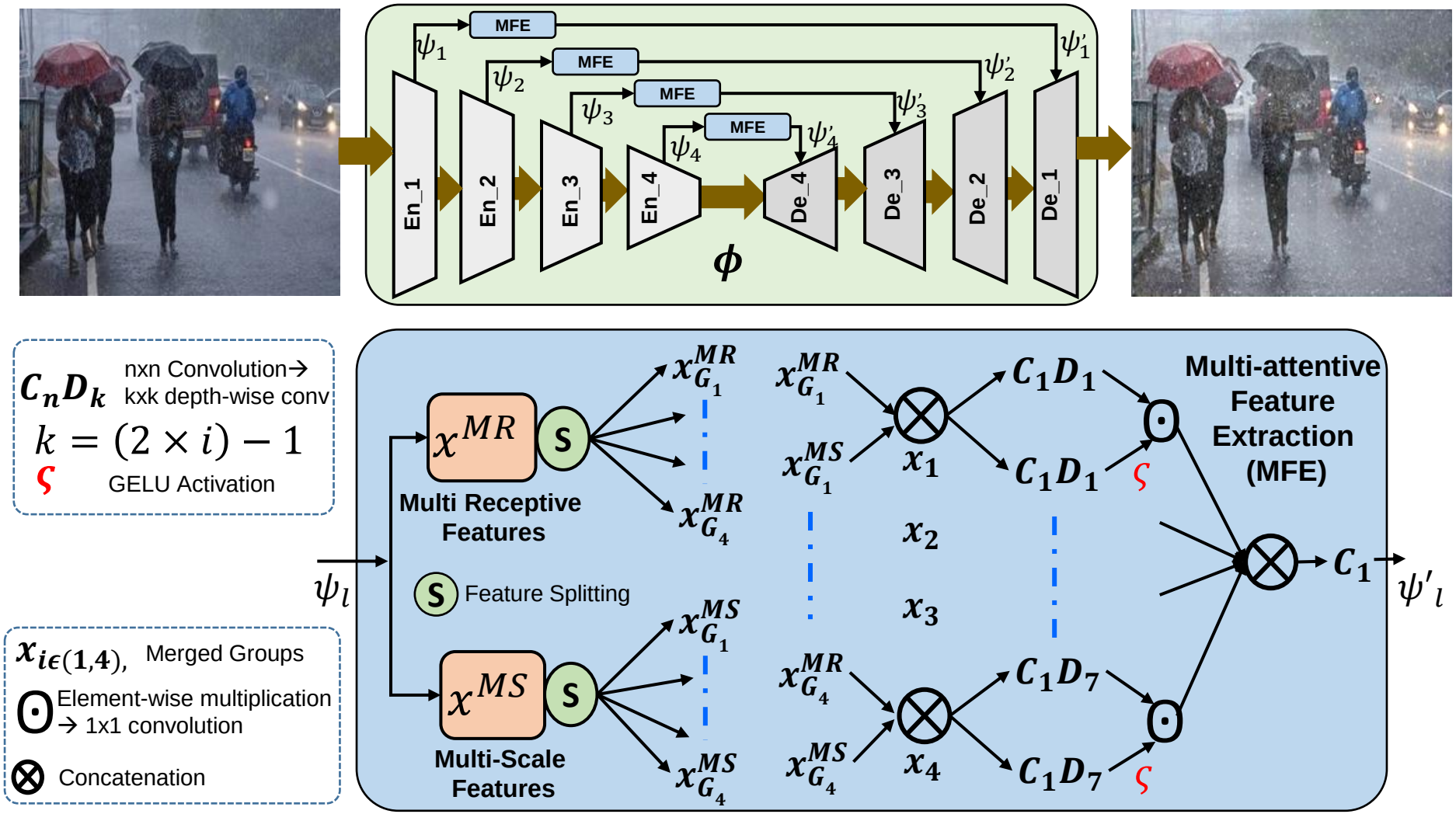
- A novel unified architecture for multi-weather image restoration is proposed. It utilizes a single trainable model to restore all weather degradations.
- A domain-translation with multi-attentive feature learning to achieve unified model's generalizability on various real-world degradations.
- A progressive multi-domain deformable alignment with cascaded multi-head attention modules are proposed for multi-weather image restoration.

# Proposed Solution (ICCV-2023)



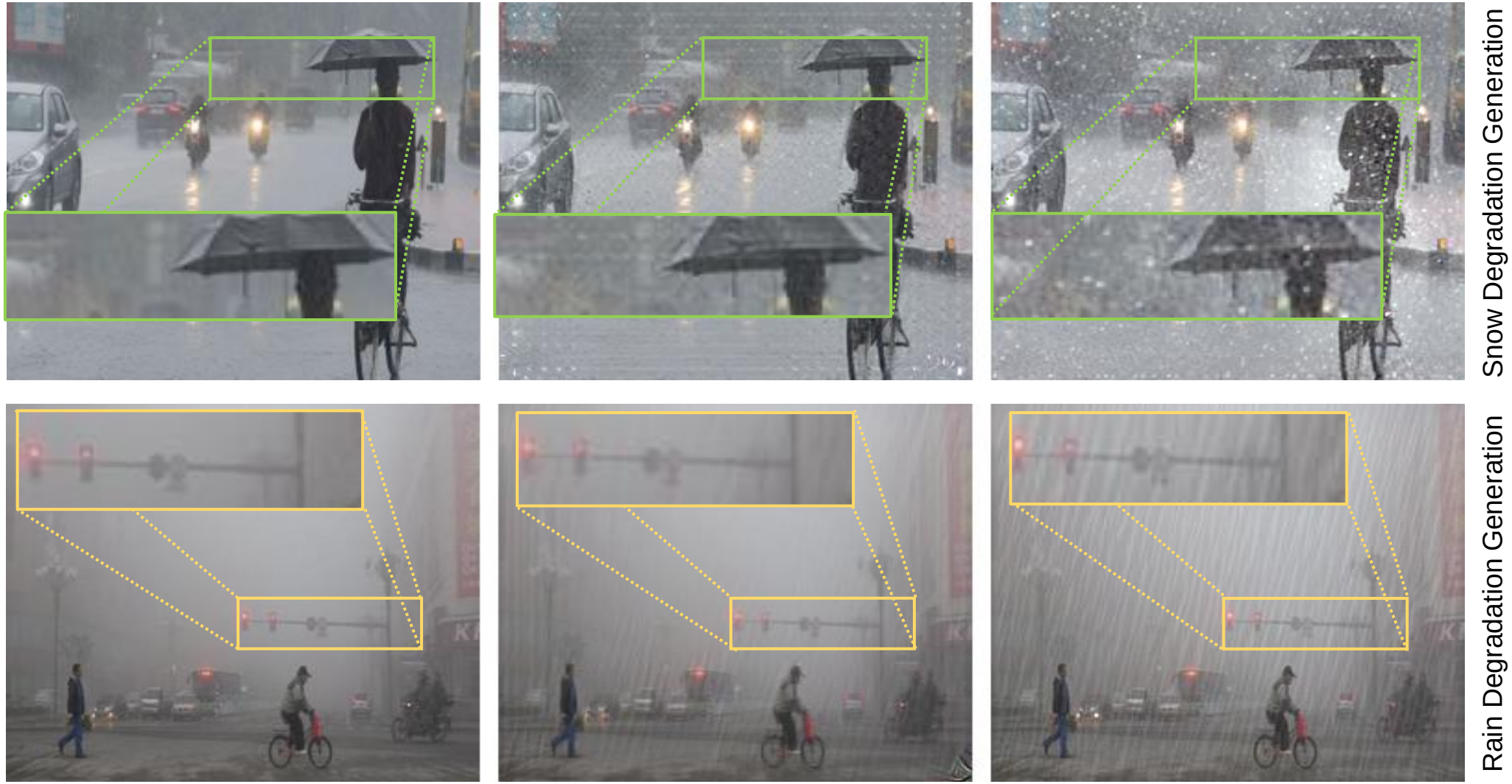

**Overview of the proposed multi-weather image restoration algorithm. Initially, the weather degraded image is translated to different domains. Next, these translated and input images are processed through separate encoders and aligned using proposed progressive multi-domain feature alignment with cascaded multi-head attention module**

# Proposed Domain Translation

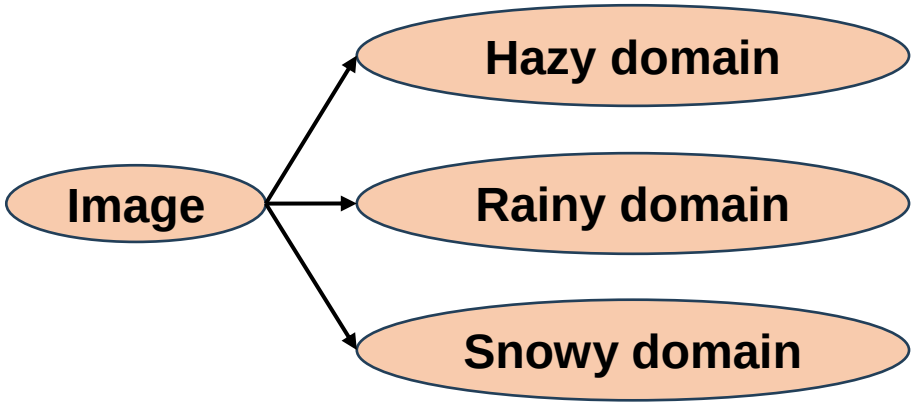


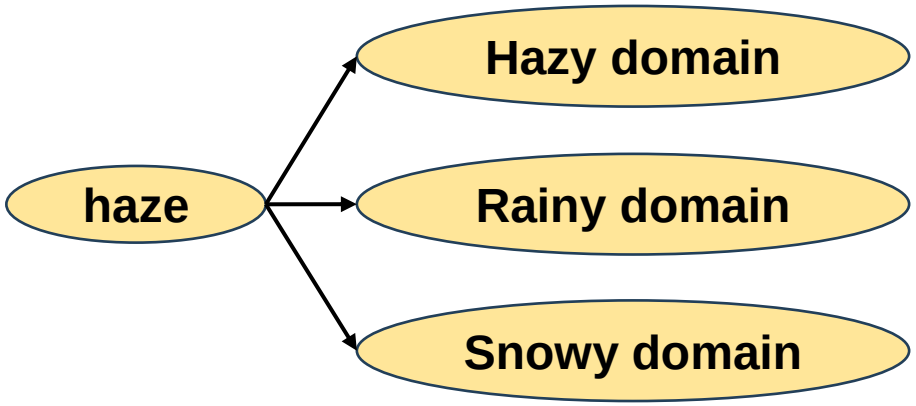

**Overview of the proposed domain translation architecture with multi-attentive feature extraction (MFE) module.**

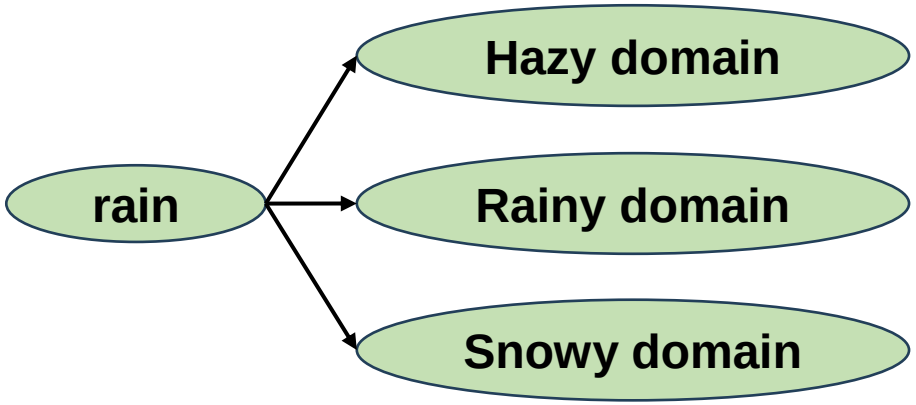
# Ablation with and without MFE in DT

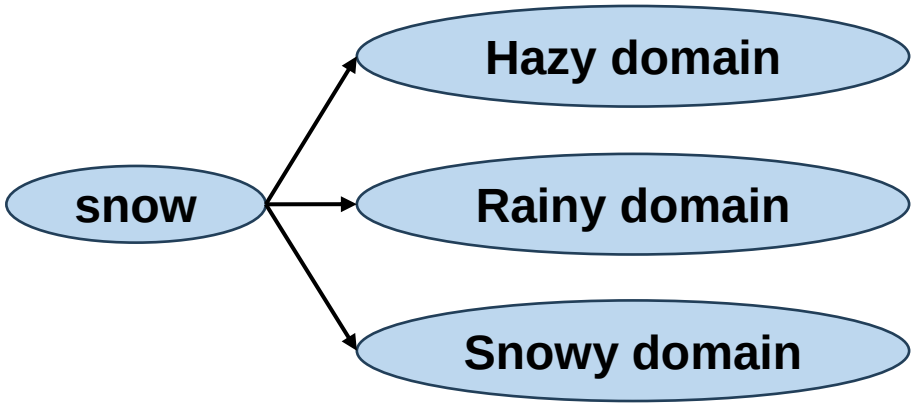


# Proposed Domain Translation



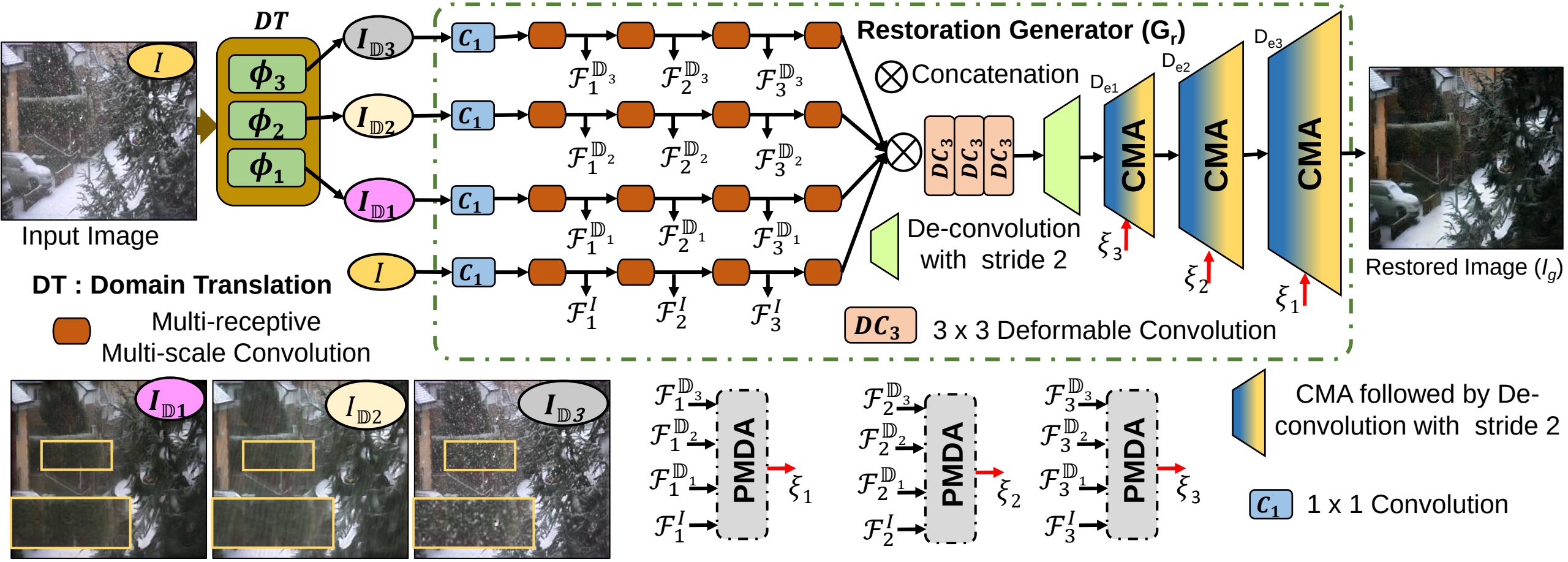



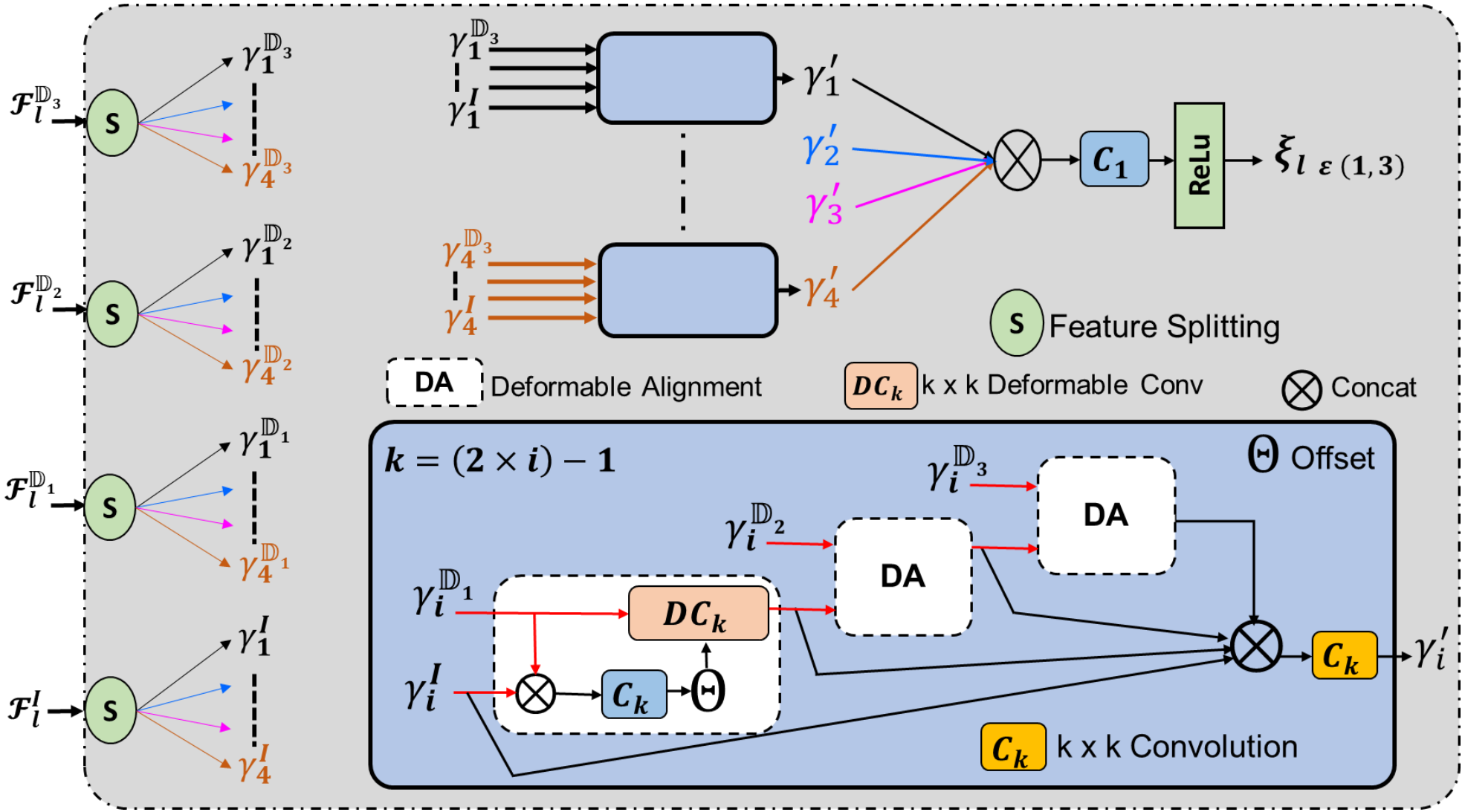
# Sample Results of Domain Translation

# Proposed Solution (ICCV-2023)




**Overview of the proposed multi-weather image restoration algorithm. Initially, the weather degraded image is translated to different domains. Next, these translated and input images are processed through separate encoders and aligned using proposed progressive multi-domain feature alignment with cascaded multi-head attention module**

# Proposed Restoration Modules




**Overview of the proposed progressive multi-domain deformable alignment module for effective feature aggregation**

# 2D Convolution Cont…

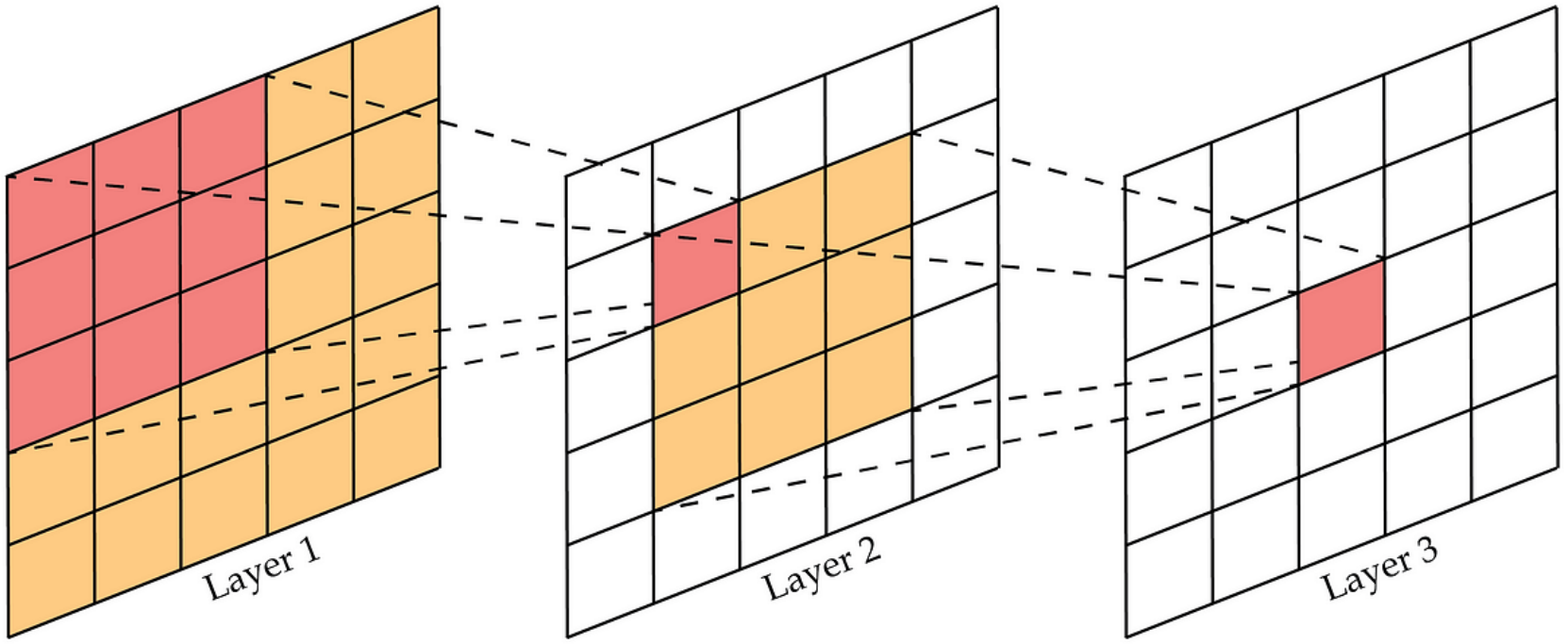
## Receptive field

- The region in the input space that a particular CNN's feature is affected by.

# 2D Convolution Cont…

## Receptive field

- The region in the input space that a particular CNN's feature is affected by.

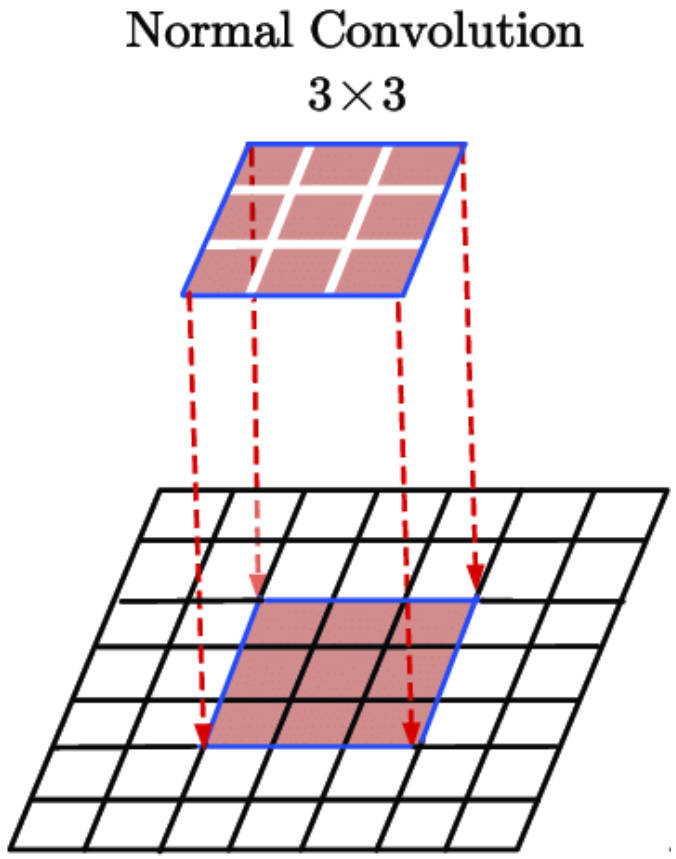


# 2D Convolution Cont…

## Dilated Convolution

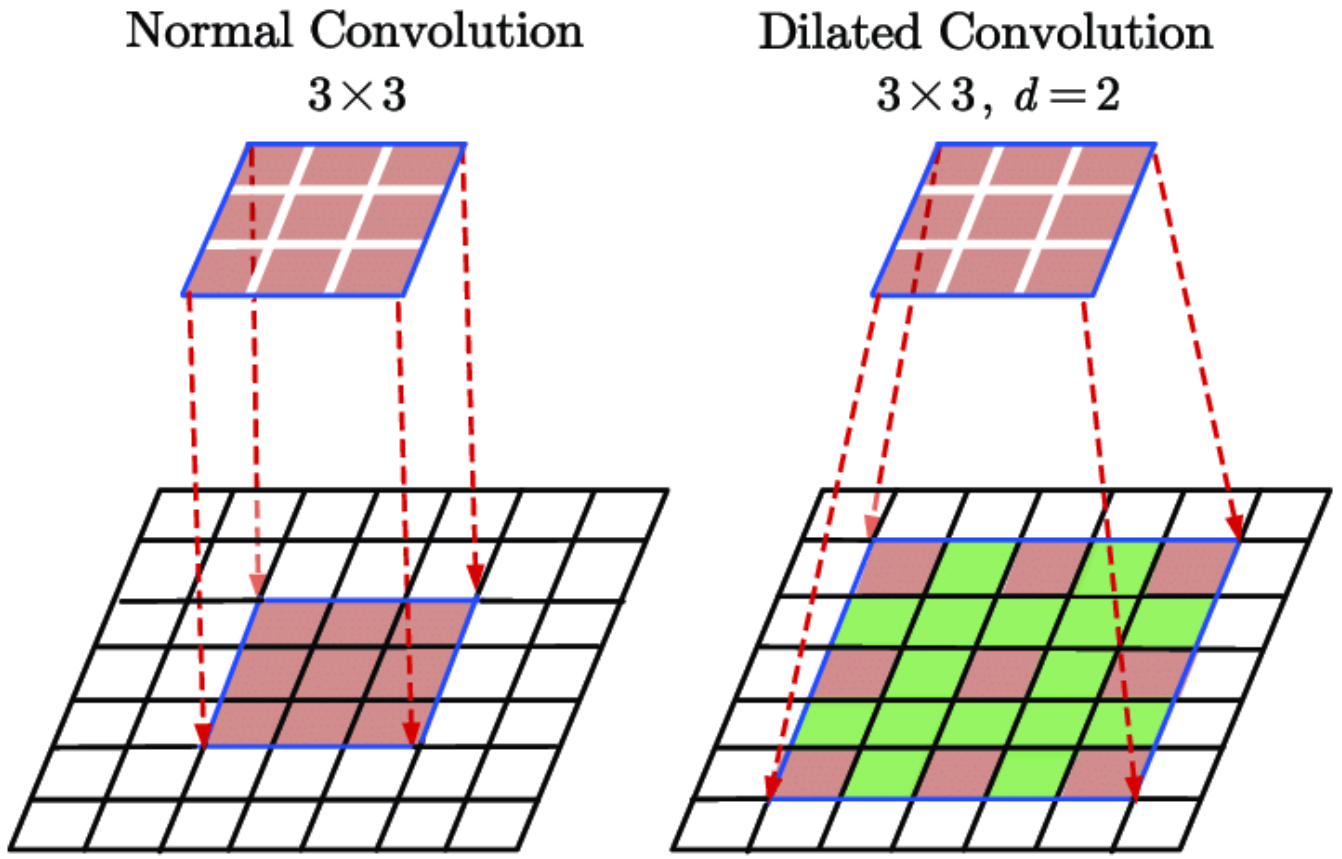
# 2D Convolution Cont…

## Dilated Convolution



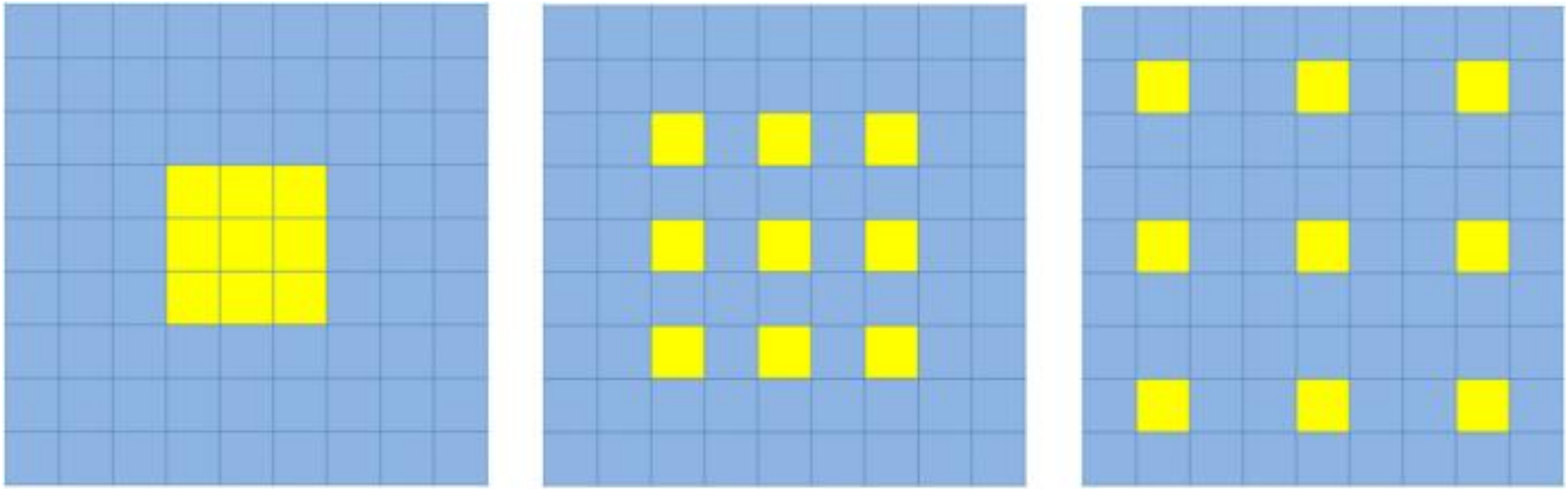
# 2D Convolution Cont…

## Dilated Convolution



# 2D Convolution Cont…

## Dilated Convolution



Kernel size： 3 × 3
Dilation rate: 1

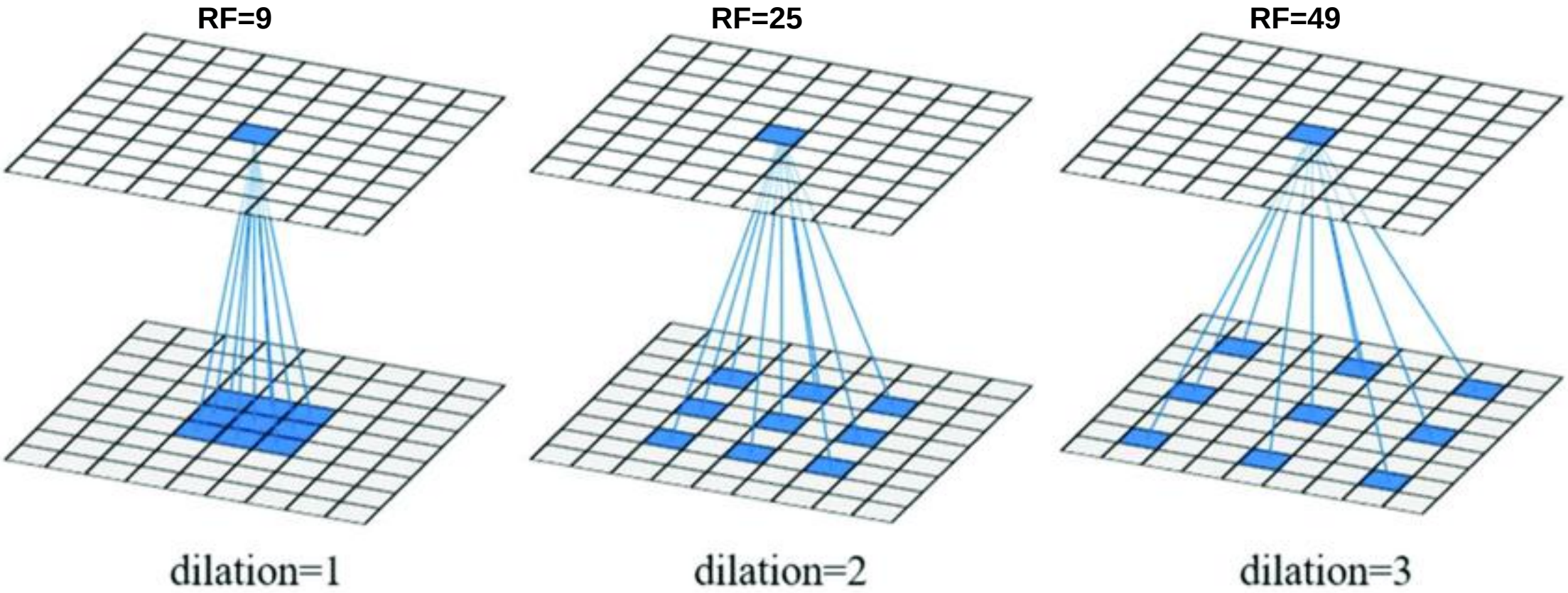
Kernel size： 3 × 3
Dilation rate: 2

Kernel size： 3 × 3
Dilation rate: 3

# 2D Convolution Cont…

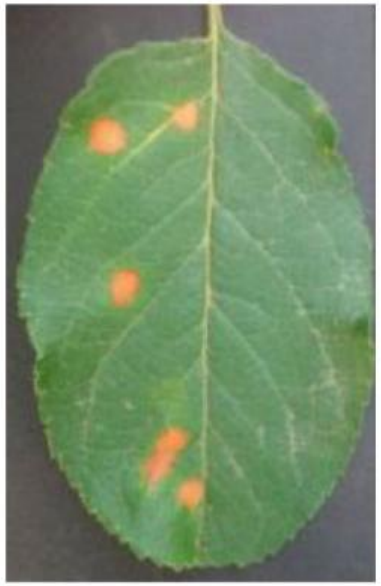
## Dilated Convolution

- Receptive Field w.r.t. previous layer (RF)

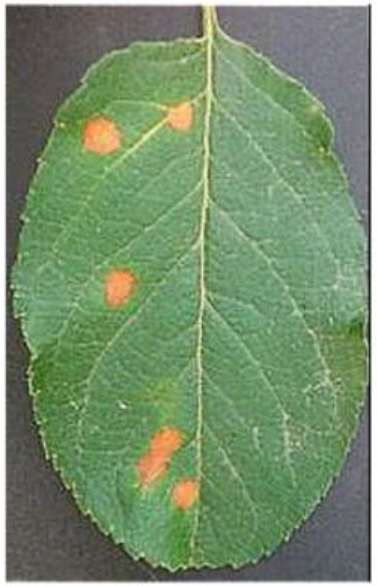


# 2D Convolution Cont…

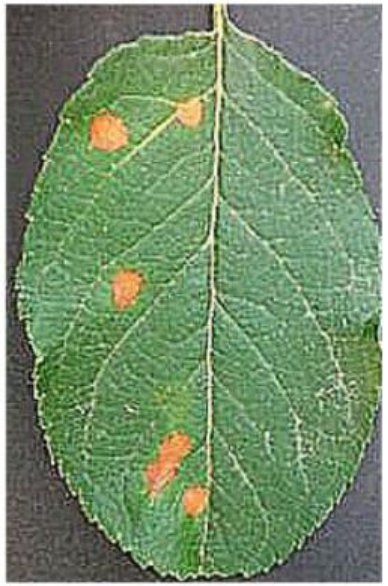
## Dilated Convolution



D Original image



E Dilated rate=1



F Dilated rate=2

G Dilated rate=3

H Dilated rate=5

Application: Apple leaf disease identification

# 2D Convolution Cont…

- Dilated convolution remains a valuable tool in the deep learning toolbox, particularly for tasks where capturing spatial context over larger receptive fields is critical

# 2D Convolution Cont…

- Dilated convolution remains a valuable tool in the deep learning toolbox, particularly for tasks where capturing spatial context over larger receptive fields is critical

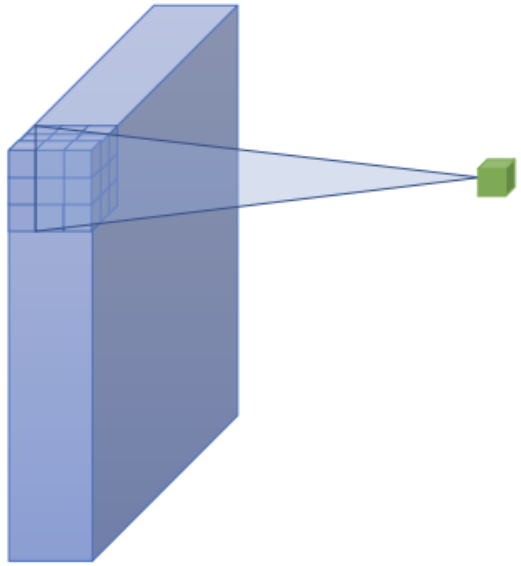
## Limitations

- **Loss of Local Detail**: with larger dilation rates can lead to a loss of local detail in the feature maps.

- **Limited Adaptability to Input Variability:** It operates on fixed-size receptive fields determined by the dilation rate and kernel size. This may not be adaptable to inputs of varying scales, limiting the model's ability to generalize across diverse input distributions.

# Depth-wise Separable Convolution

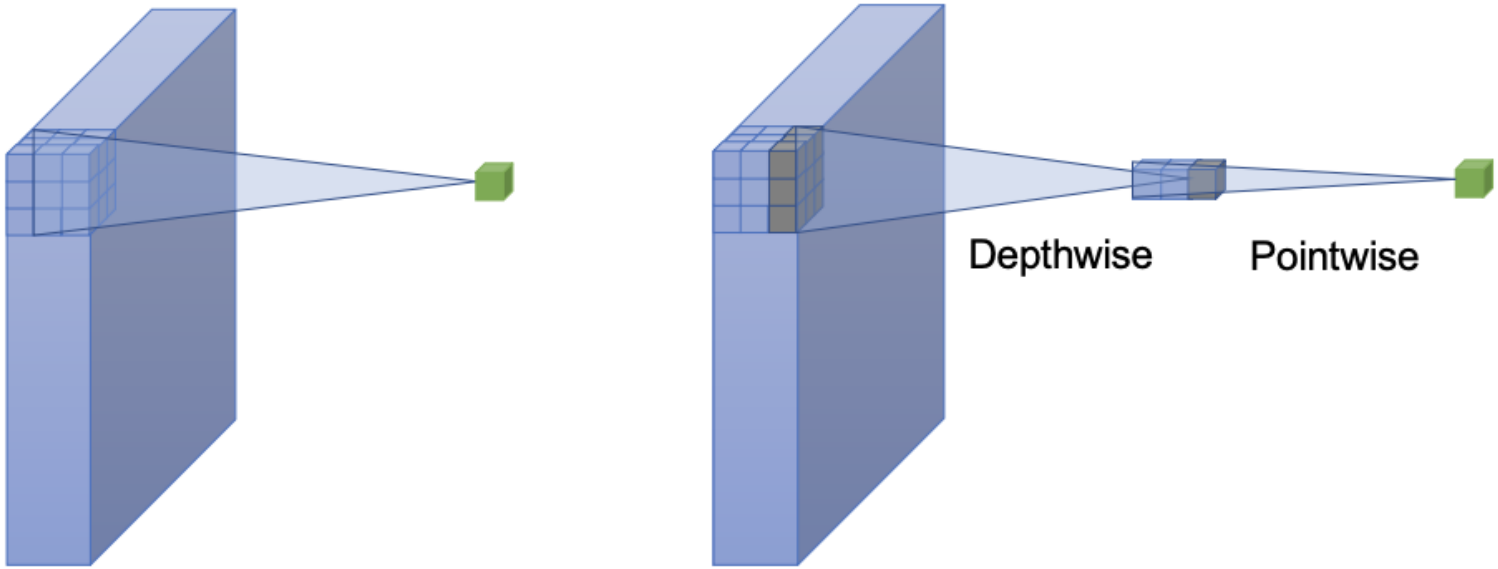
- Convolution performs the channel-wise and spatial-wise computation in one step.



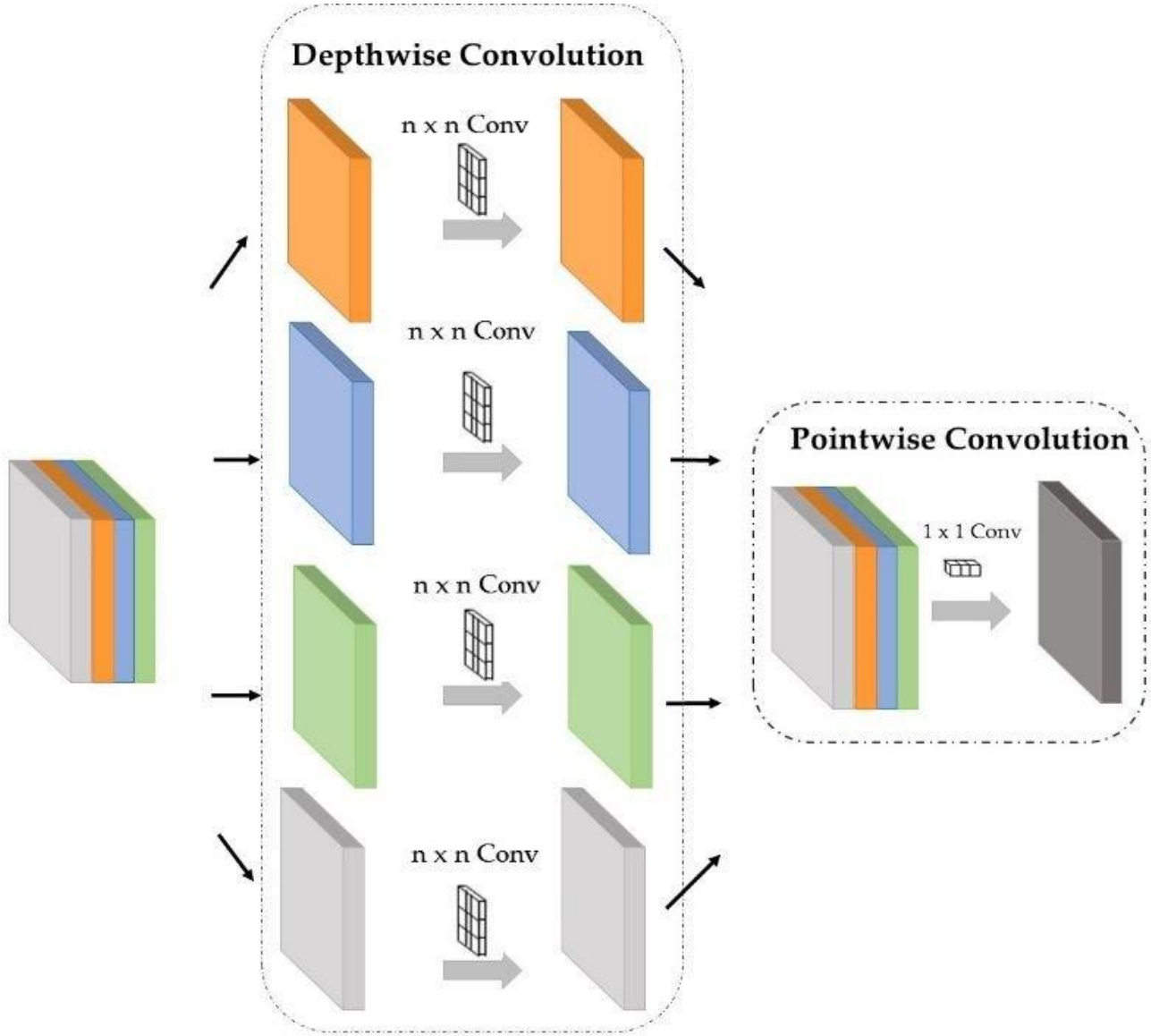
# Depth-wise Separable Convolution

- Convolution performs the channel-wise and spatial-wise computation in one step.

- Depth-wise Separable Convolution splits the computation into two steps:
  - a single convolutional filter per each input channel
  - pointwise convolution is used to create a final feature maps.



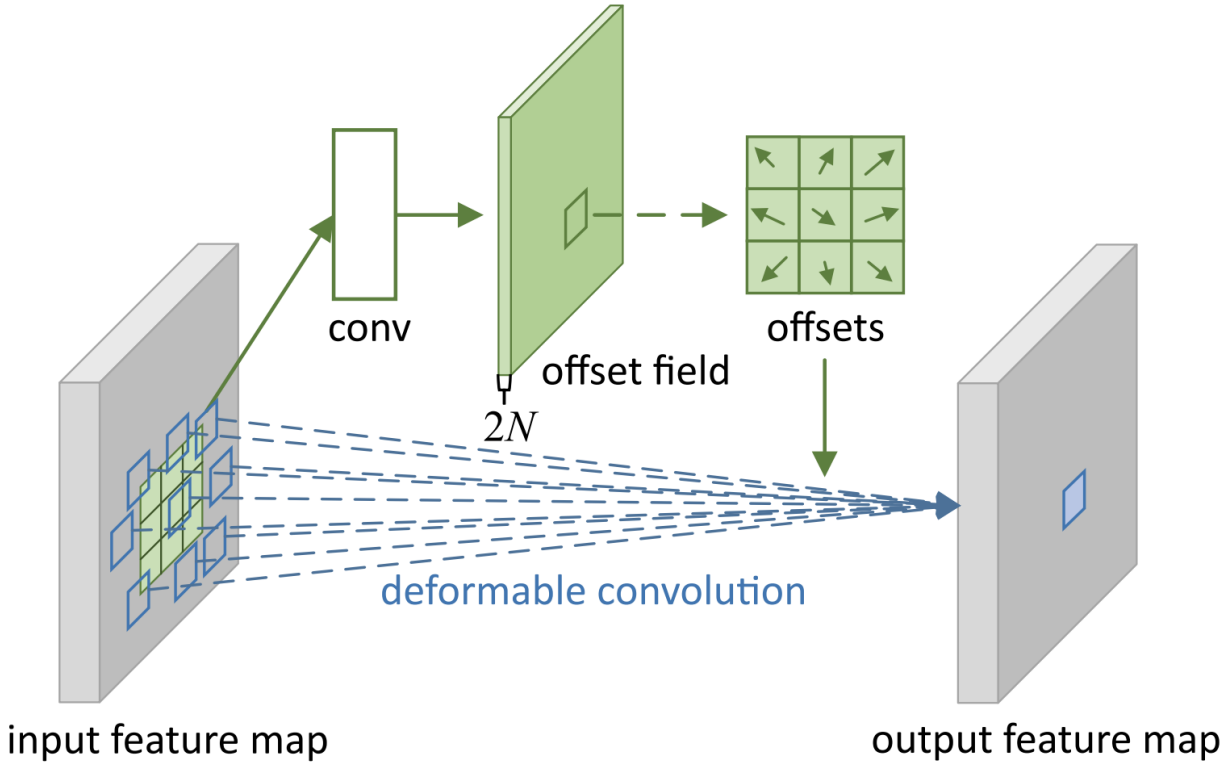
# Depth-wise Separable Convolution



# Deformable Convolution

- an advanced convolution technique in deep learning designed to enhance the flexibility of standard convolutional operations.

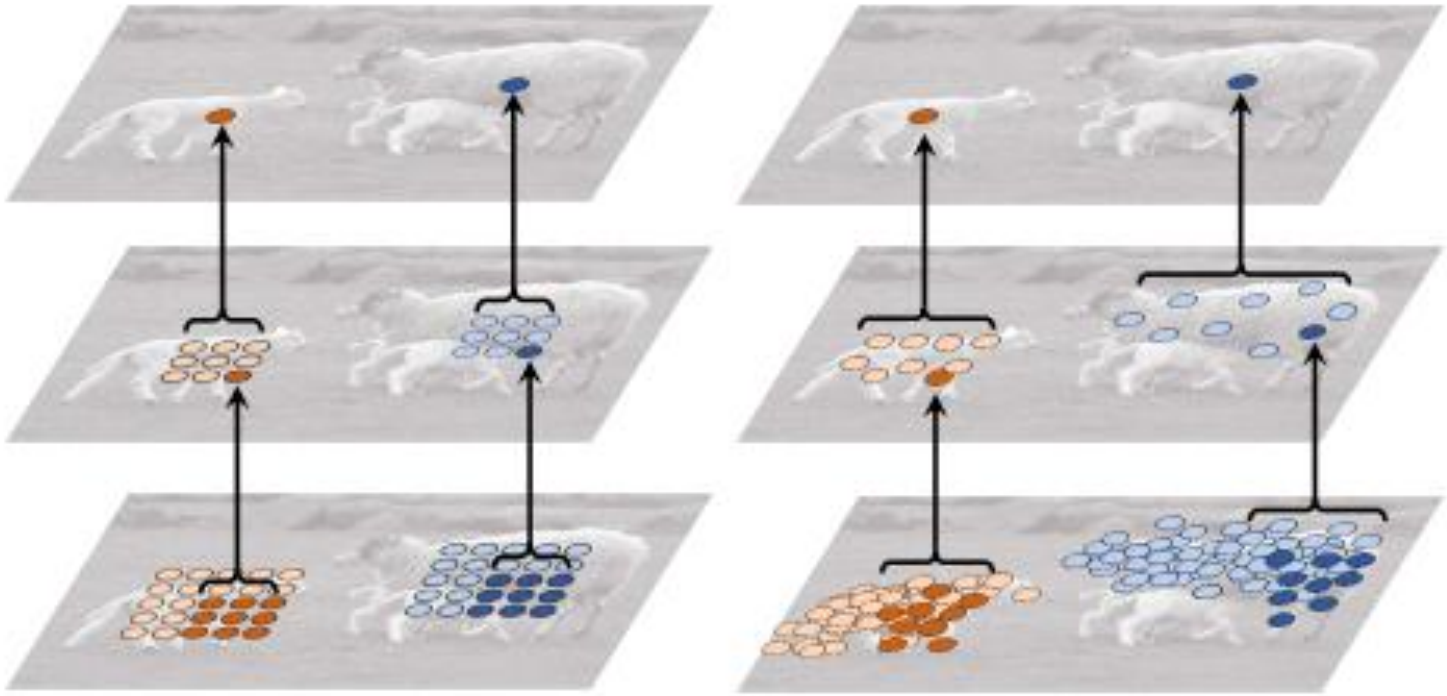
- Deformable convolution introduces learnable offsets to the sampling grid of a convolution kernel



# Deformable Convolution

- an advanced convolution technique in deep learning designed to enhance the flexibility of standard convolutional operations.

- Deformable convolution introduces learnable offsets to the sampling grid of a convolution kernel



# Deformable Convolution

- an advanced convolution technique in deep learning designed to enhance the flexibility of standard convolutional operations.

- Deformable convolution introduces learnable offsets to the sampling grid of a convolution kernel




# Ablation on Alignment Orders

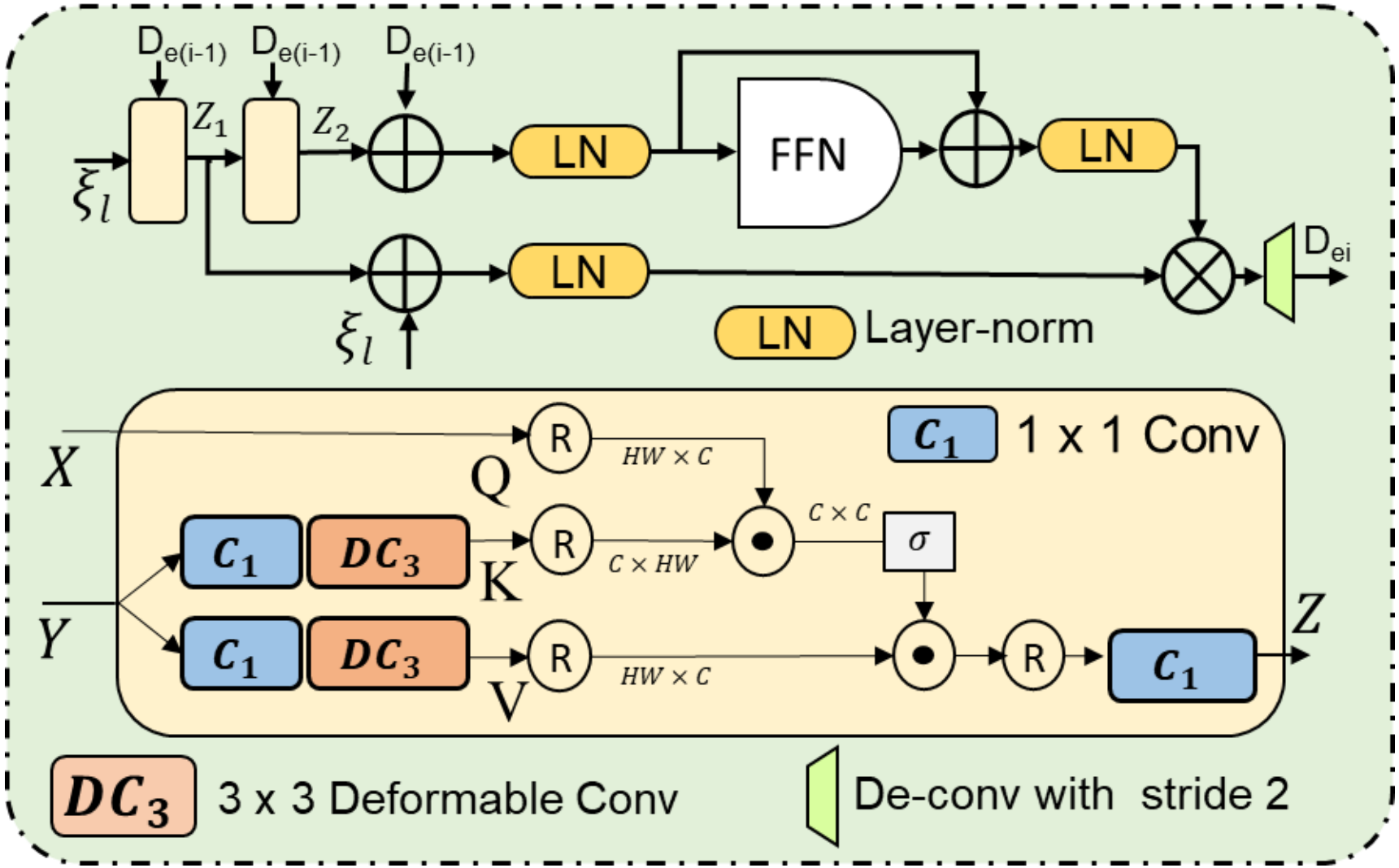
Input Image

SHR

RSH

HRS

# Proposed Restoration Modules




**Overview of the proposed cascaded multi-head attention module for feature merging.**

# Training Details

## Domain Translation:

- Domain translation network is trained separately for each considered domain like haze, rain with veil, snow with veil.

- 7500 paired images (**any degraded image and required domain**) are generated synthetically for training each domain separately.

## Restoration :

- SOTS, Outdoor rain and CSD databases are used combinedly with domain translation for training of the proposed restoration network.

- Along with synthetic database test splits, three real world degraded databases are used for experimental analysis.

# Ablation Study Analysis

**Ablation study on the proposed modules with outdoor rain database (IL: Independent Learning, PMDA: Progressive Multidomain Alignment and CMA: Cascaded Multi-head Attention)**

| Database → | ORD | | SOTS | | CSD | |
|---|---|---|---|---|---|---|
| Module ↓ | PSNR | SSIM | PSNR | SSIM | PSNR | SSIM |
| Baseline | 26.49 | 0.887 | 28.63 | 0.921 | 28.32 | 0.879 |
| Baseline + IL | 27.89 | 0.905 | 33.23 | 0.934 | 29.01 | 0.899 |
| Baseline + IL + PMDA | 29.53 | 0.942 | 35.46 | 0.978 | 31.76 | 0.932 |
| **Baseline + IL + PMDA + CMA** | **31.24** | **0.951** | **36.26** | **0.987** | **32.95** | **0.940** |

# Quantitative Results

**Subjective results on real-world weather degraded images for de-hazing, de-raining with veil and snow with veil removal.**

| Database | Methods | Pub-Year | NIQE | Entropy | BRISQUE |
|---|---|---|---|---|---|
| RTTS | UMVR | TMM-22 | 5.009 | 7.221 | 28.625 |
| | KD | CVPR-22 | 4.996 | 7.297 | 26.837 |
| | TW | CVPR-22 | 5.703 | 7.263 | 29.874 |
| | **Ours** | **ICCV-23** | **4.859** | **7.505** | **27.761** |
| RID | UMVR | TMM-22 | 6.604 | 7.486 | 24.296 |
| | KD | CVPR-22 | 6.943 | 7.459 | 24.841 |
| | TW | CVPR-22 | 7.496 | 7.393 | 24.165 |
| | **Ours** | **ICCV-23** | **7.625** | **7.492** | **23.931** |
| Snow_Realistic | UMVR | TMM-22 | 4.296 | 7.354 | 23.761 |
| | KD | CVPR-22 | 6.643 | 7.459 | 24.841 |
| | TW | CVPR-22 | 5.628 | 7.331 | 25.377 |
| | **Ours** | **ICCV-23** | **4.196** | **7.575** | **21.884** |

# Result Analysis

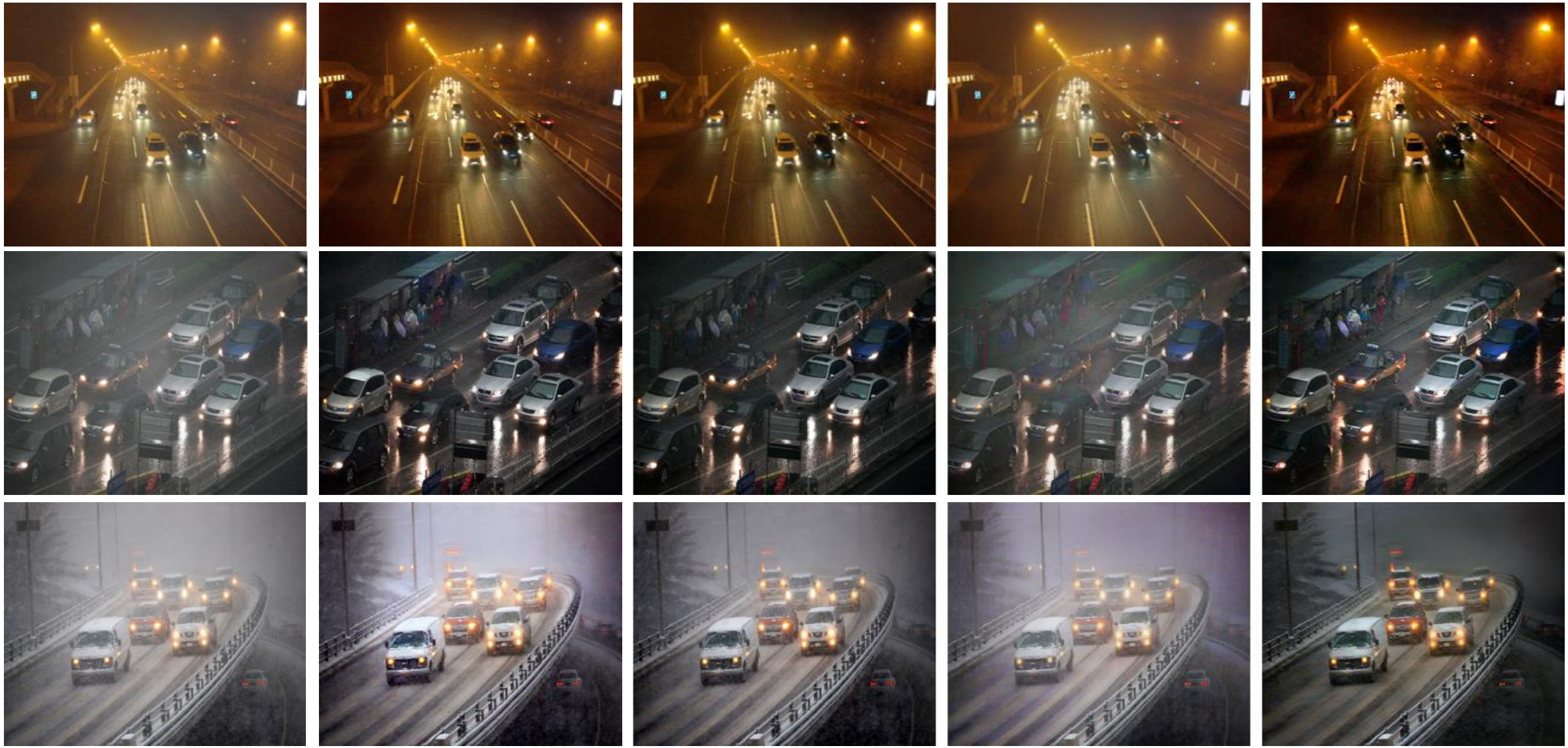
**Quantitative results analysis for de-hazing (SOTS), deraining with veil (ORD) and snow with veil (CSD) removal in terms of average PSNR/SSIM.**

| Methods | SOTS | CSD | ORD |
|---|---|---|---|
| UMVR [1] | 33.41/0.980 | 28.65/0.900 | 22.99/0.830 |
| KD [2] | 34.64/0.985 | 31.35/**0.950** | 29.05/0.916 |
| TW [3] | 32.45/0.955 | 29.76/0.940 | 27.96/0.950 |
| Ours | **36.26/0.987** | **32.95**/0.942 | **31.24/0.951** |

[1]. Kulkarni et al., “Unified multi-weather visibility restoration”, IEEE Transactions on Multimedia, 2022
[2]. Chen et al., “Learning multiple adverse weather removal via two-stage knowledge learning and multi-contrastive regularization: Toward a unified model”, CVPR, 2022
[3]. Valanarasu et al., “Transweather: Transformer-based restoration of images degraded by adverse weather conditions”, CVPR 2022

# Visual Results

# Cross-scene Visual Results

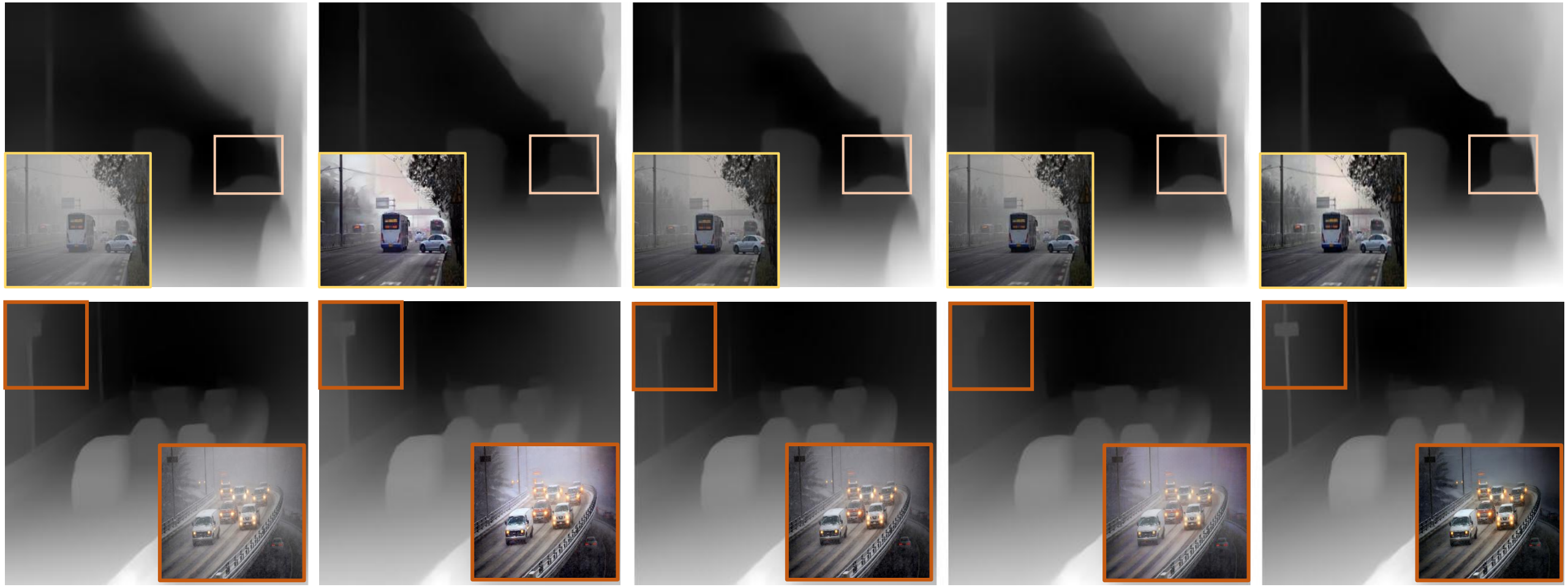


Input Image | UMVR | KD | TW | Ours

# Applications of the Proposed Work


Input Image | UMVR | KD | TW | Ours

# Applications of the Proposed Work



Input Image | UMVR | KD | TW | Ours

# Conclusion

- First weather-invariant representation learning based multi-domain progressive deformable alignment architecture for multi-weather image restoration.
- Weather-invariant representation learning achieved via domain translation to learn a representation that is common across diverse weather conditions.
- Proposed progressive multi-domain deformable alignment module effectively merges the independently learned multi-domain features with effective restoration via proposed CMA module.
- Results analysis is conducted on real-world and synthetic benchmark hazy, rainy and snowy datasets. Also, we have analysed the effectiveness of the proposed and existing methods for various tasks.

# Limitations of Current Video Restoration

- Effective algorithm for multi-weather degraded image restoration.
- Limited unified multi-weather video restoration algorithm.
- Efficient temporally consistent algorithm for video restoration.
- Databases for day-night time snow removal, dust removal.
- All weather activity recognition, anomaly detection, object tracking algorithms.

# Thank you

**For more details,**

**Please visit: https://github.com/pwp1208**