# Exploring the Up-sampling Filter for Video applications (EVC-UFV) V.1.0 standard

Online, 2025/07/23T13 UTC

Associate Professor - Alessandro Artusi

CYENS CENTRE OF EXCELLENCE

Deep Camera

MPAI. community

# Who is MPAI

- MPAI – Moving Picture, Audio and Data Coding by Artificial Intelligence is an international, unaffiliates, non-profit association based in Geneva.
- MPAI mission is the promotion of the efficient use of Data by
  - Developing Technical Specifications for
    - Coding of any type of Data, especially using new technologies such as Artificial Intelligence, and
    - Technologies that facilitate integration of Data Coding and Decoding components in Information and Communication Technology systems, and by
  - Bridging the gap between Technical Specifications and their practical use through the development of Intellectual Property Rights Guidelines (“IPR Guidelines”),.
- MPAI operates based on open international collaboration of interested parties supporting the MPAI mission and the means to accomplish it.

# MPAI Standards

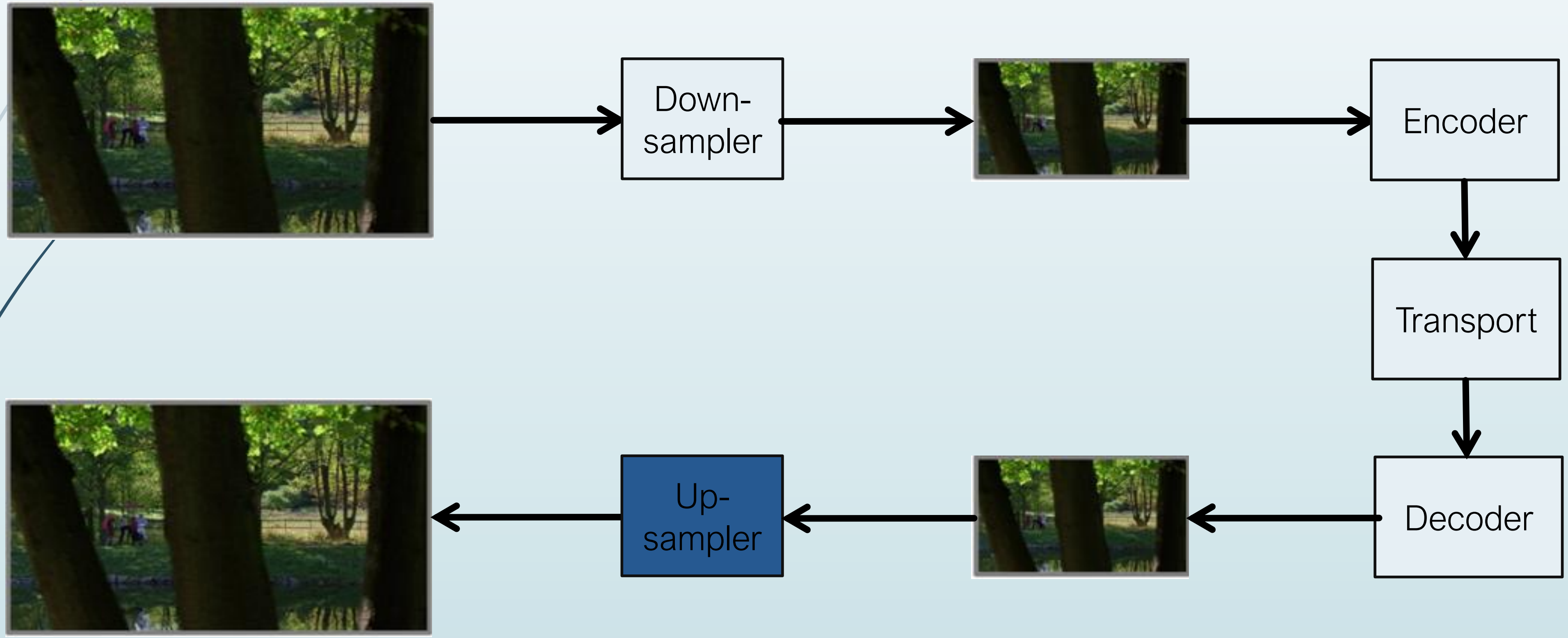
## The latest Standard

- AI Framework
- Audio Enhancement
- Automotive
- Video Up-sampling Filter
- Finance
- Gaming
- Governance
- Human-Machine Communication
- Multimodal Conversation
- Metaverse
- Neural Network Traceability
- Neural Network Watermarking
- Objects and Scenes
- Portable Avatar
- Profiles
- Qualifiers

MPAI. community

# Framework

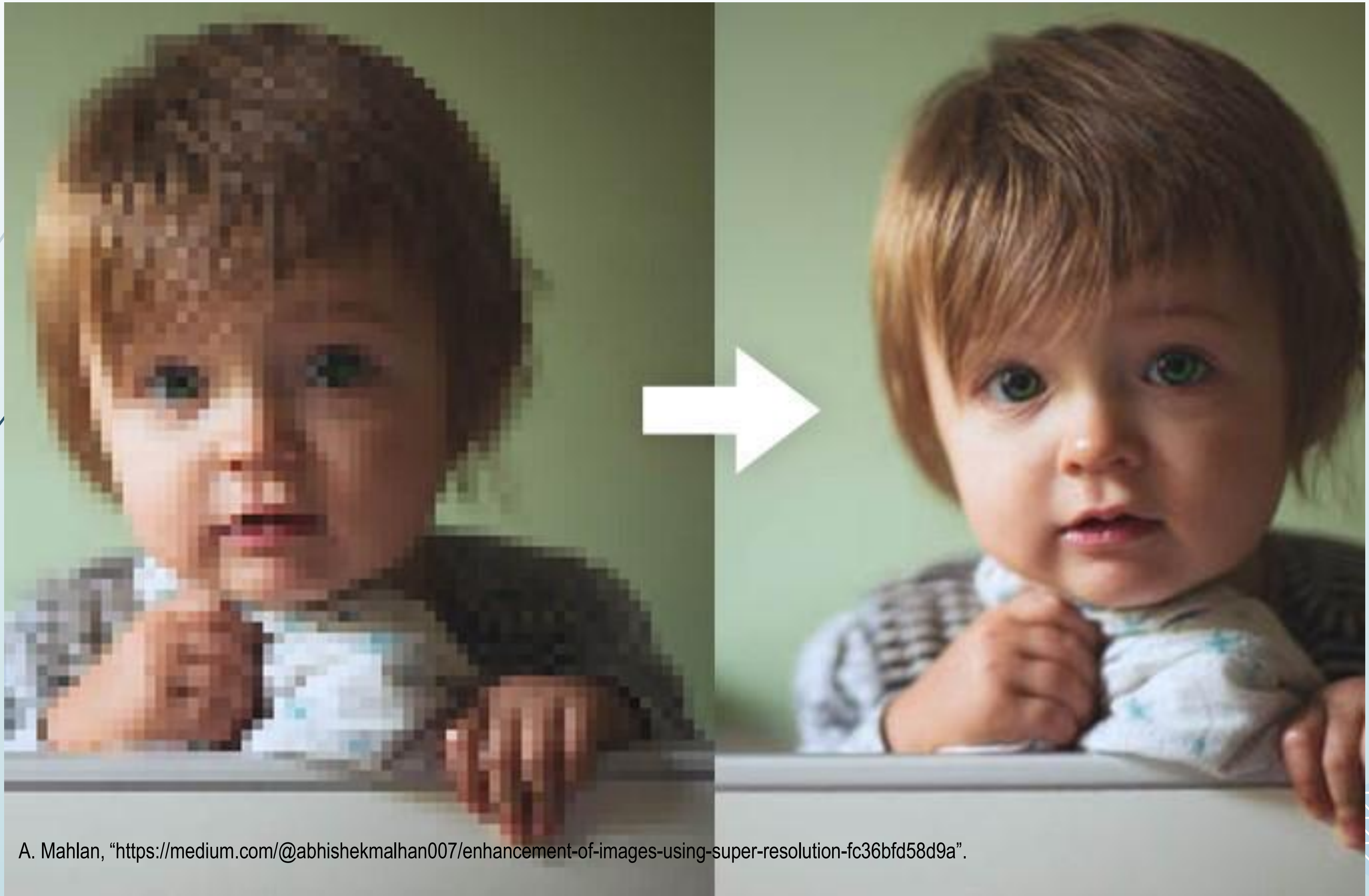
# Up-sampling for video coding






MPAI.
community

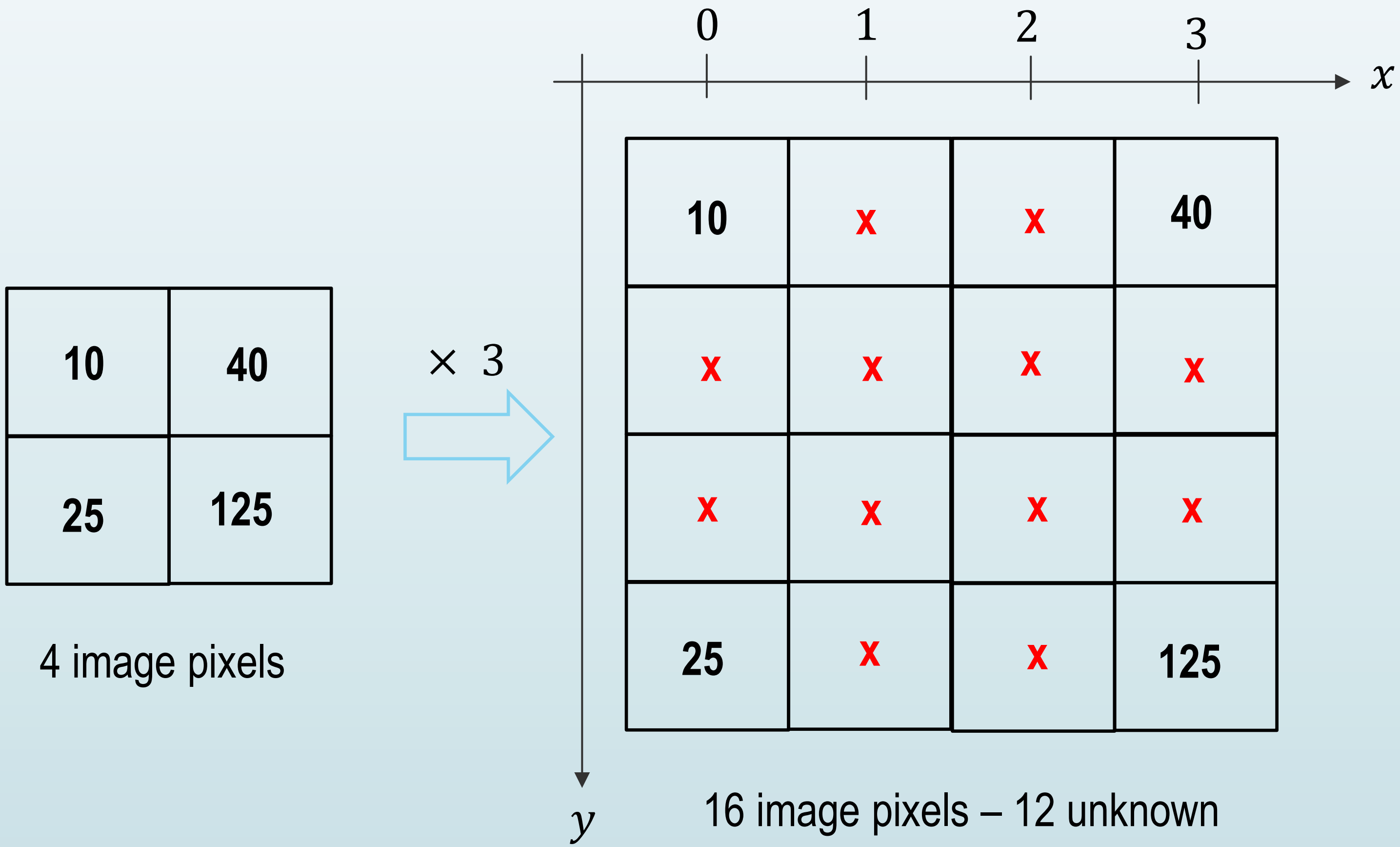
# Super resolution: What is it?



A. Mahlan, “https://medium.com/@abhishekmalhan007/enhancement-of-images-using-super-resolution-fc36bfd58d9a”.

MPAI. community

# Super resolution: What is it?



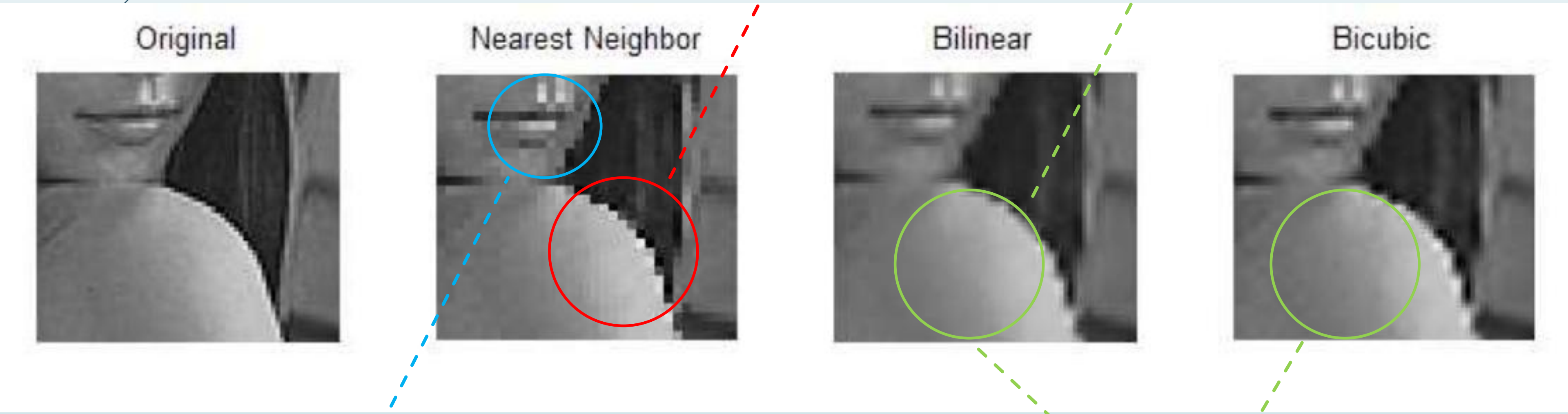

4 image pixels

16 image pixels – 12 unknown

# Traditional techniques

Aliasing artifacts along edges

Blurring artifacts




Pixelation artifacts

Lost of details

# Deep-learning based techniques

Ground Truth image

Bicubic patch

Ground Truth patch

Deep-learning patch

MPAI.
community

# Up-sampling filter requirements


SD
540 pixels
960 pixels
× 2
HD
1080 pixels
1920 pixels
HD
× 2
4K
2160 pixels
3840 pixels

# Requirements

The technology should be agnostic to the decoding system

| | |
|---|---|
| **Bits/sample** | 8 and 10 bit-depth per component. |
| **Encoding technology** | HEVC, VVC and AVC. |
| **Colous space** | YCbCr with 4:2:0 sub-sampling. |
| **Encoding settings** | Random Access and Low Delay at QPs 22, 27, 32, 37, 42, 47. |
| **Up-sampling supported** | SD to HD - HD to 4K. |
| **Complexity** | Low complexity. |
| | |

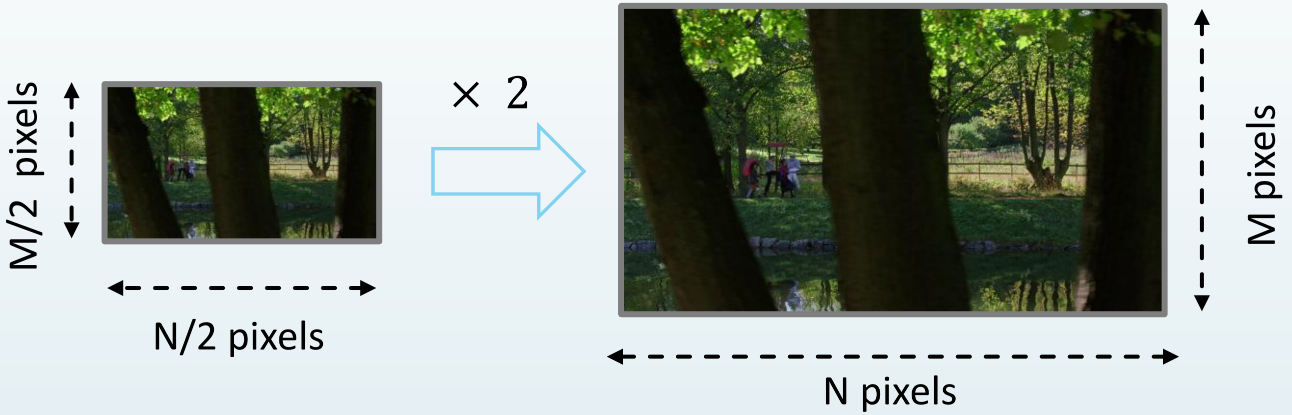
The technology should support up-sampling by 2 and 4

MPAI. community

# 1. Design Procedure

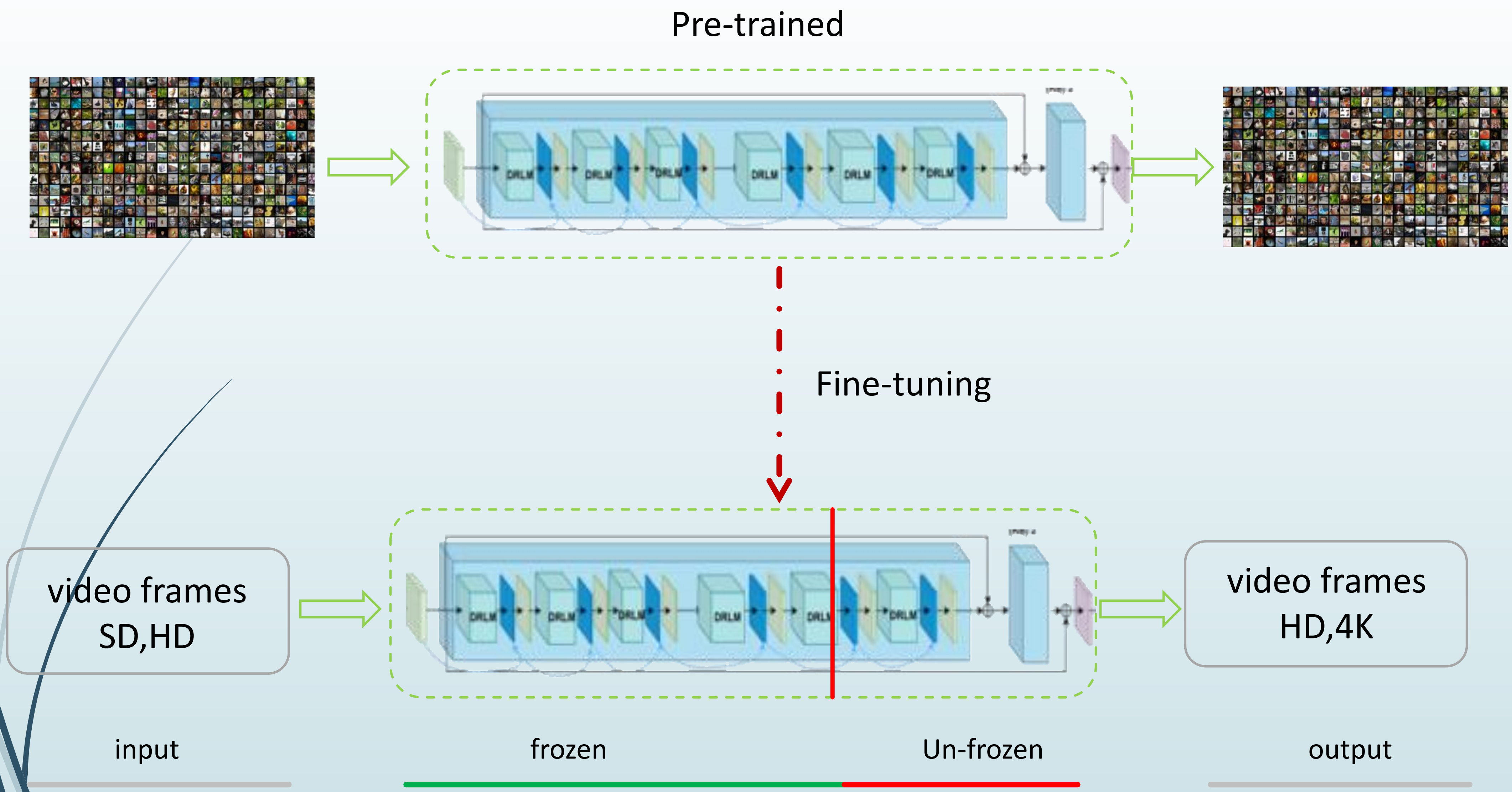
# 1.1 Data Preparation

# 1. Dataset: pair of images




# 2. Dataset: Images decomposed in patches

# 1.2 Training


Pre-trained
Fine-tuning
video frames
SD,HD
video frames
HD,4K
input
frozen
Un-frozen
output

# 1.3 Network Design

# Baseline: Densely Residual Laplacian Network/1

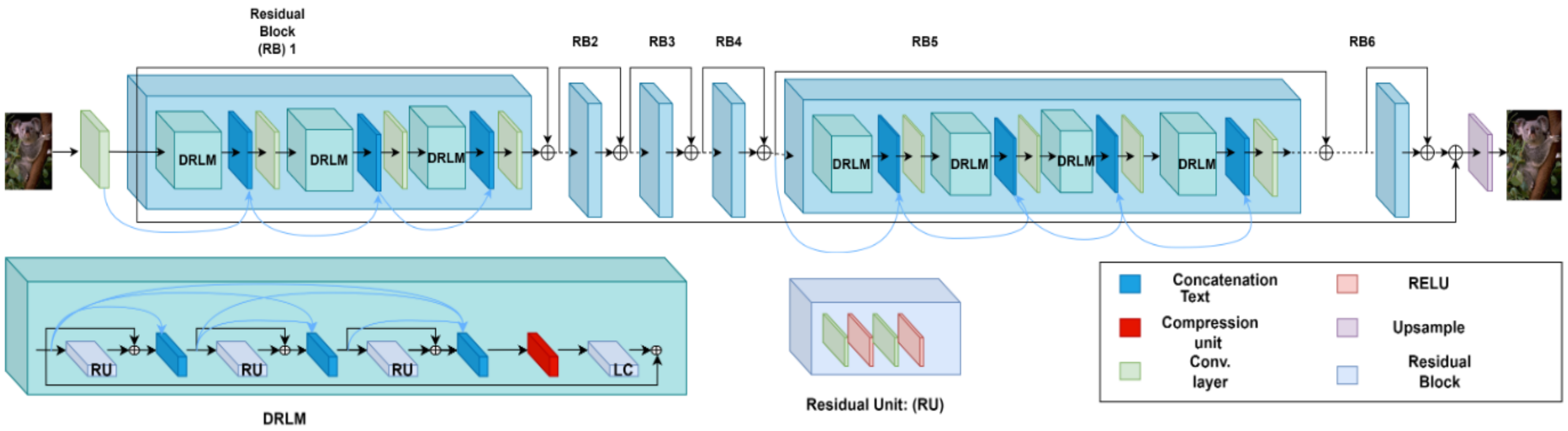



**Fig. 1. The detailed network architecture of the baseline model.** The top figure shows the whole network architecture consisting of six cascaded residual blocks (RB). The bottom figure shows the internal structure of sub-components i.e. densely residual laplacian module (DRLM) and Residual Units(RU).

S. Anwar and N.Barnes, "Densely Residual Laplacian Super-resolution", IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI), 2020.

# Baseline: Densely Residual Laplacian Network/2

6 Residual Blocks (RB)

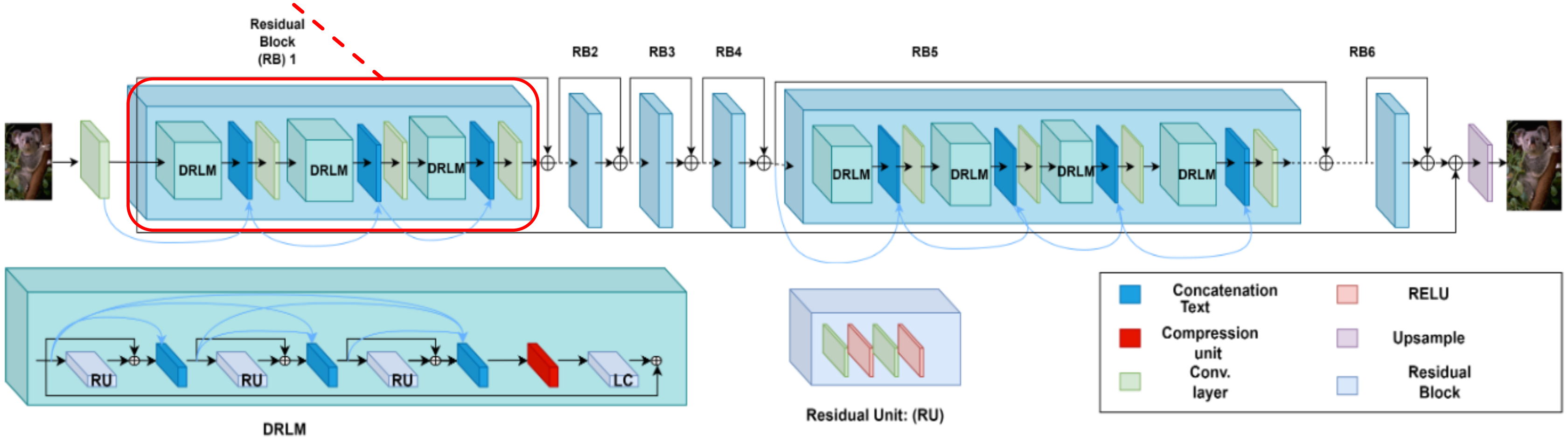



**Fig. 1. The detailed network architecture of the baseline model.** The top figure shows the whole network architecture consisting of six cascaded residual blocks (RB). The bottom figure shows the internal structure of sub-components i.e. densely residual laplacian module (DRLM) and Residual Units(RU).

S. Anwar and N.Barnes, "Densely Residual Laplacian Super-resolution", IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI), 2020.

# Baseline: Densely Residual Laplacian Network/3

6 Residual Blocks (RB)

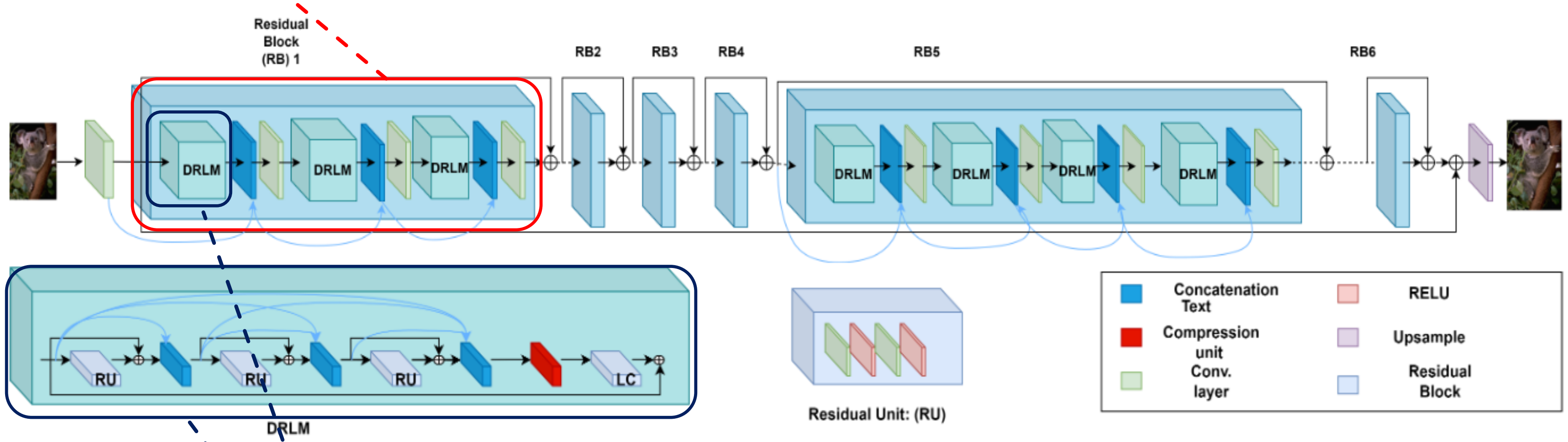



**Fig. 1. The detailed network architecture of the baseline model.** The top figure shows the whole network architecture consisting of six cascaded residual blocks (RB). The bottom figure shows the internal structure of sub-components i.e. densely residual laplacian module (DRLM) and Residual Units(RU).

20 Densely Residual Laplacian Modules (DRLM)

S. Anwar and N.Barnes, "Densely Residual Laplacian Super-resolution", IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI), 2020.

# Baseline: Densely Residual Laplacian Network/4

6 Residual Blocks (RB)

3x(20 DRLM) Residual Unit (RU), 60 in total

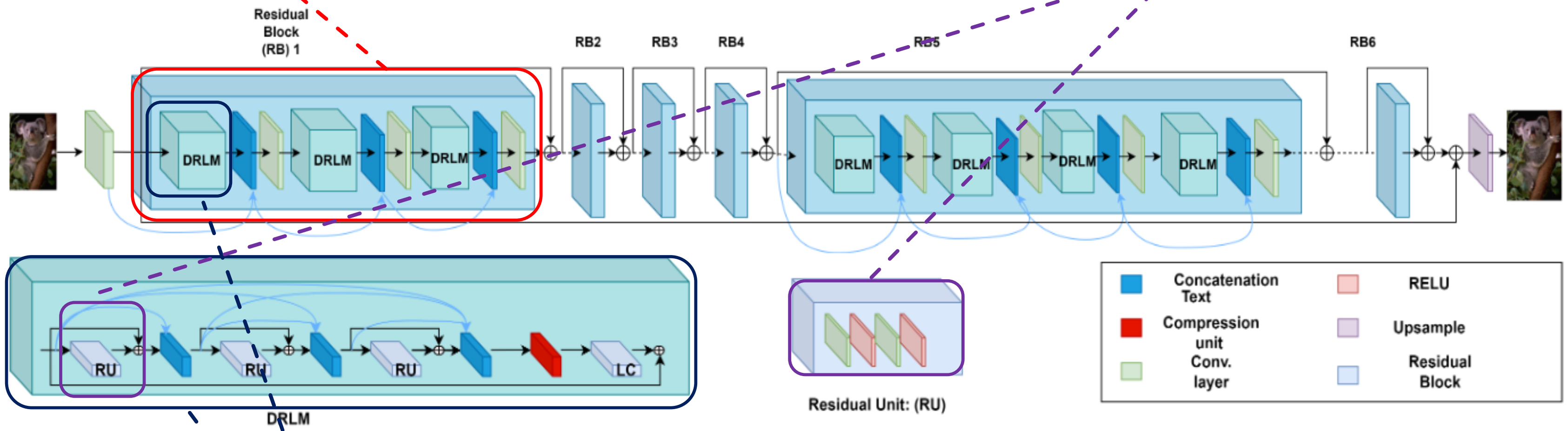



**Fig. 1**. **The detailed network architecture of the baseline model**. The top figure shows the whole network architecture consisting of six cascaded residual blocks (RB). The bottom figure shows the internal structure of sub-components i.e. densely residual laplacian module (DRLM) and Residual Units(RU).

20 Densely Residual Laplacian Modules (DRLM)

S. Anwar and N.Barnes, "Densely Residual Laplacian Super-resolution", IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI), 2020.

# Baseline:  Densely Residual Laplacian Network/5

6 Residual Blocks (RB)

3x(20 DRLM) Residual Unit (RU), 60 in total

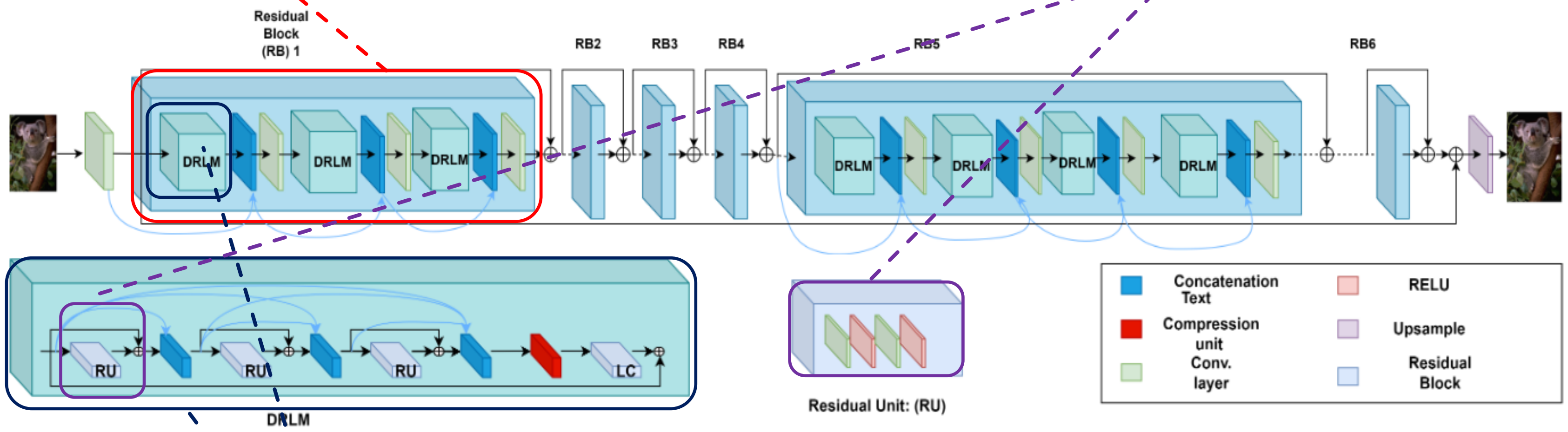



**Fig. 1. The detailed network architecture of the baseline model.** The top figure shows the whole network architecture consisting of six cascaded residual blocks (RB). The bottom figure shows the internal structure of sub-components i.e. densely residual laplacian module (DRLM) and Residual Units(RU).

20 Densely Residual Laplacian Modules (DRLM)

Lower-level components:

2x(60 RU) Hidden layers (HL), 120 in total

64 Features Maps (FM) per conv. layer

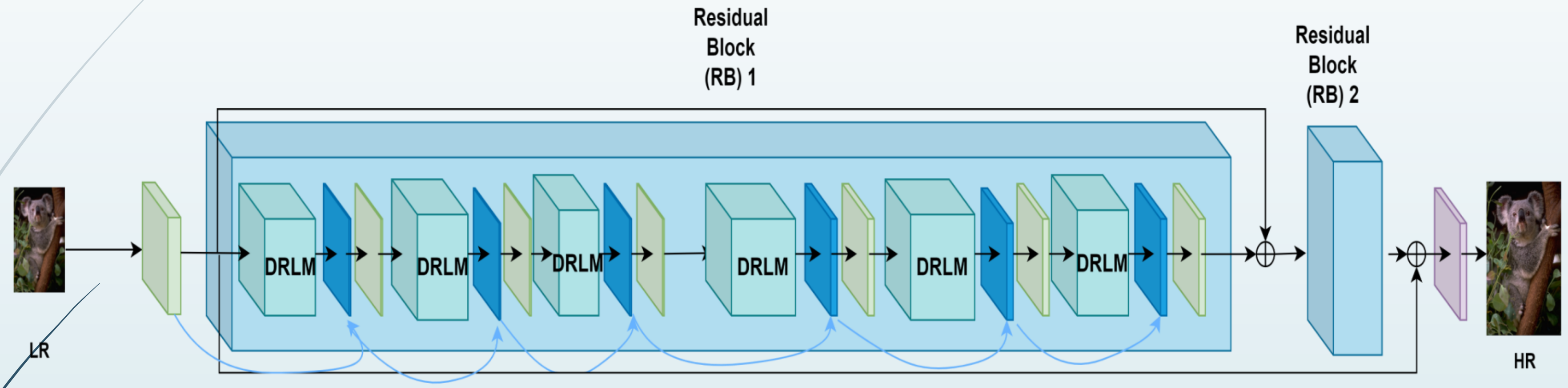
S. Anwar and  N.Barnes, "Densely Residual Laplacian Super-resolution", IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI), 2020.

# Comparison and model selection

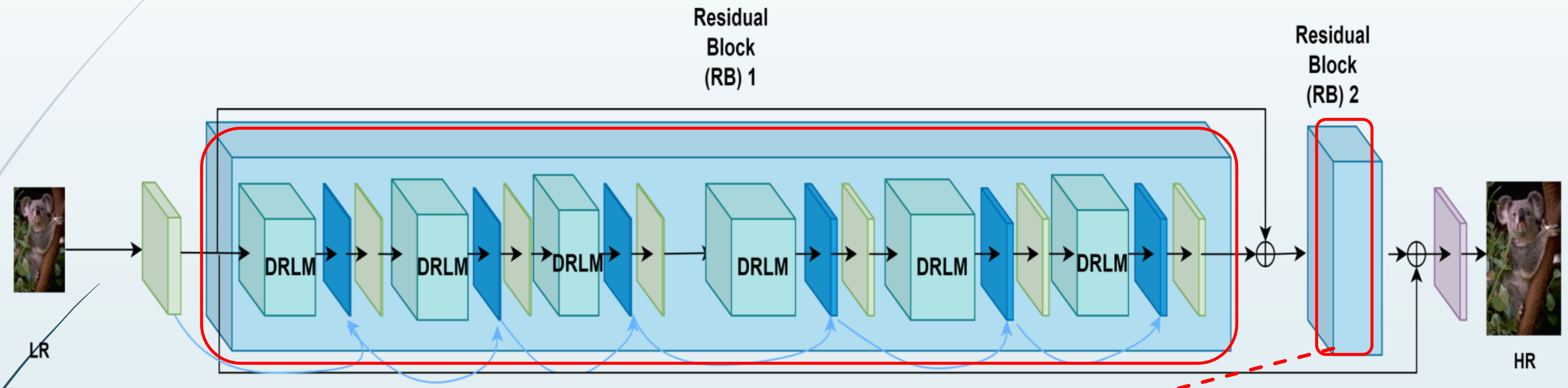
| Model | Structural Changes | | | | | PSNR score(dB) | | Avg. Runtime(s) | | GPU Memory(MB) | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | RB | DRLM | RU | HL | FM | U100 | B100 | U100 | B100 | Res.1 | Res.2 | Res.3 | Res.4 |
| BM | 6 | 20 | 3 | 2 | 64 | 32.603 | 32.262 | 2.83 | 1.1 | 2036 | 8035 | 12173.7 | 12173.7 |
| M1 | 6 | 20 | 3 | 2 | 32 | 32.093 | 32.138 | 2.31 | 0.98 | 992.1 | 4026 | 6119 | 6119 |
| M2 | 6 | 20 | 3 | 1 | 64 | 32.502 | 32.239 | 2.61 | 1.08 | 1182 | 4774 | 7253 | 7253 |
| M3 | 6 | 20 | 2 | 2 | 64 | 32.2 | 32.2 | 2.04 | 0.52 | 1071 | 4320 | 6562 | 6562 |
| M4 | 4 | 12 | 3 | 2 | 64 | 32.388 | 32.201 | 2.47 | 0.97 | 1233 | 4867 | 7375 | 7375 |
| M5 | 2 | 12 | 3 | 2 | 64 | 32.585 | 32.26 | 2.01 | 0.49 | 644 | 2585 | 3925 | 3925 |
| M6 | 4 | 12 | 3 | 1 | 32 | 32.1855 | 32.074 | 1.57 | 0.46 | 365.5 | 1482 | 2252 | 2252 |
| M7 | 3 | 12 | 3 | 1 | 32 | 31.867 | 32.075 | 1.61 | 0.48 | 366 | 1483 | 2253 | 2253 |
| M8 | 2 | 12 | 3 | 1 | 32 | 31.893 | 32.097 | 1.64 | 0.48 | 366.1 | 1483 | 2253 | 2253 |

Table 1. **Comparison between the proposed models (M1 to M8) and the baseline model (BM).** The first section-columns refer to structural changes in the main components of the network. The next three section-columns provide PSNR Score in dB, avg. Runtime per second, GPU memory occupation in MB on two benchmark datasets , i.e., URBAN100 [10] and B100 [11]. Here the red, green and blue color depict the changes in main components, the best performing model for the selected characteristic, and second best performing models considering the trade off of other deciding characteristics respectively.

# Model distilled/1

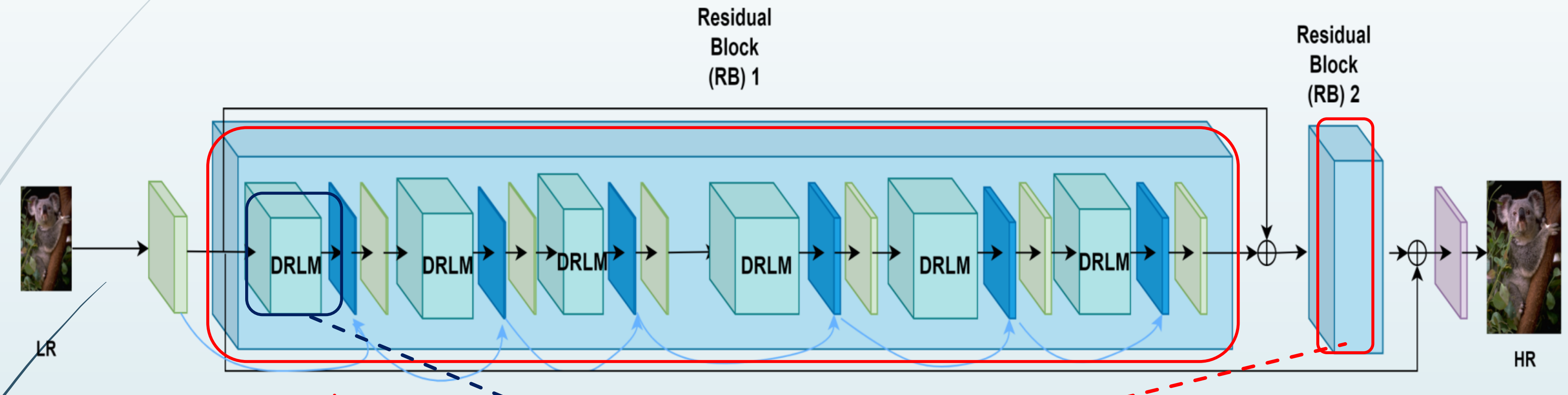


# Model distilled/2




2 Residual Blocks (RB)

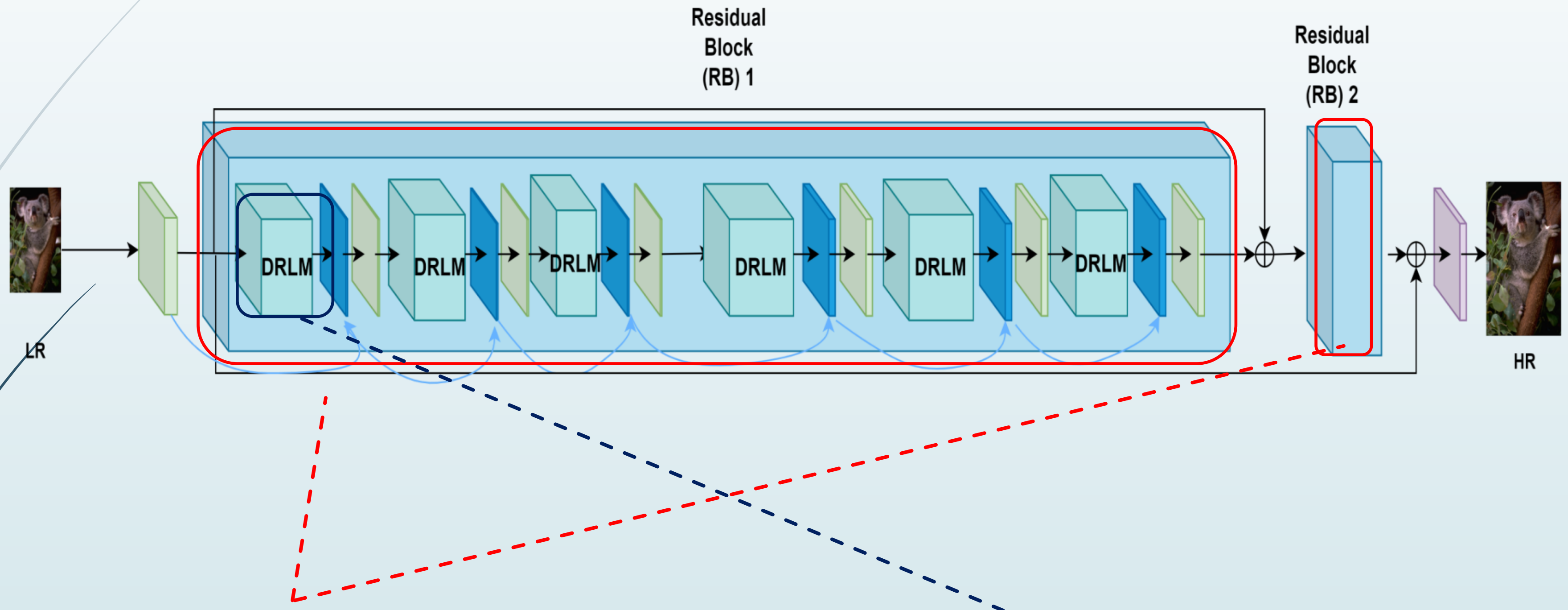
# Model distilled/3




2 Residual Blocks (RB)

6x(2 RB) Densely Residual Laplacian Modules (DRLM), 12 in total

# Model distilled/4

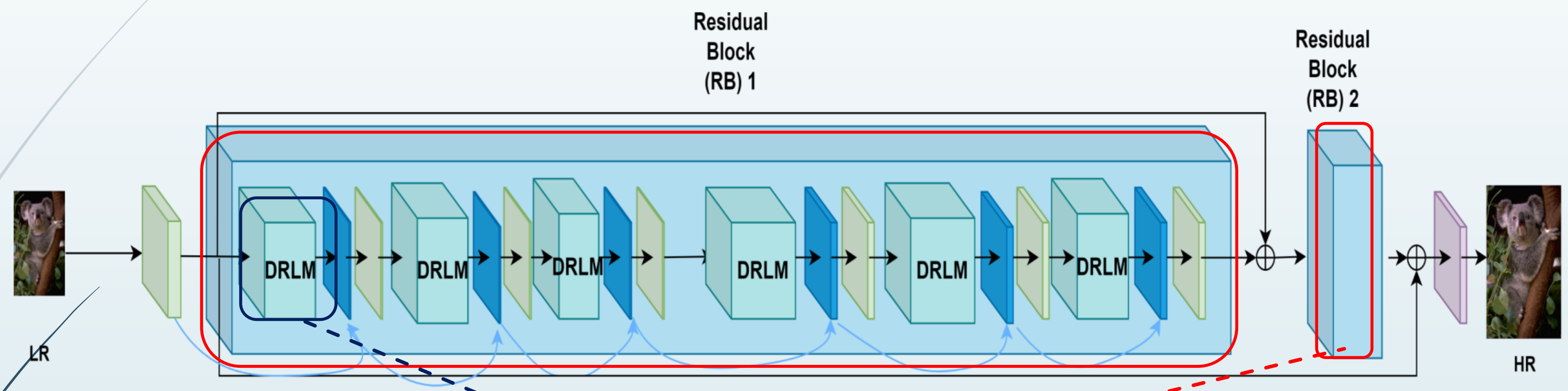



2 Residual Blocks (RB)

6x(2 RB) Densely Residual Laplacian Modules (DRLM), 12 in total

3x(12 DRLM) Residual Unit (RU), 36 in total

# Model distilled/5




2 Residual Blocks (RB)

6x(2 RB) Densely Residual Laplacian Modules (DRLM), 12 in total

3x(12 DRLM) Residual Unit (RU), 36 in total

Lower-level components:

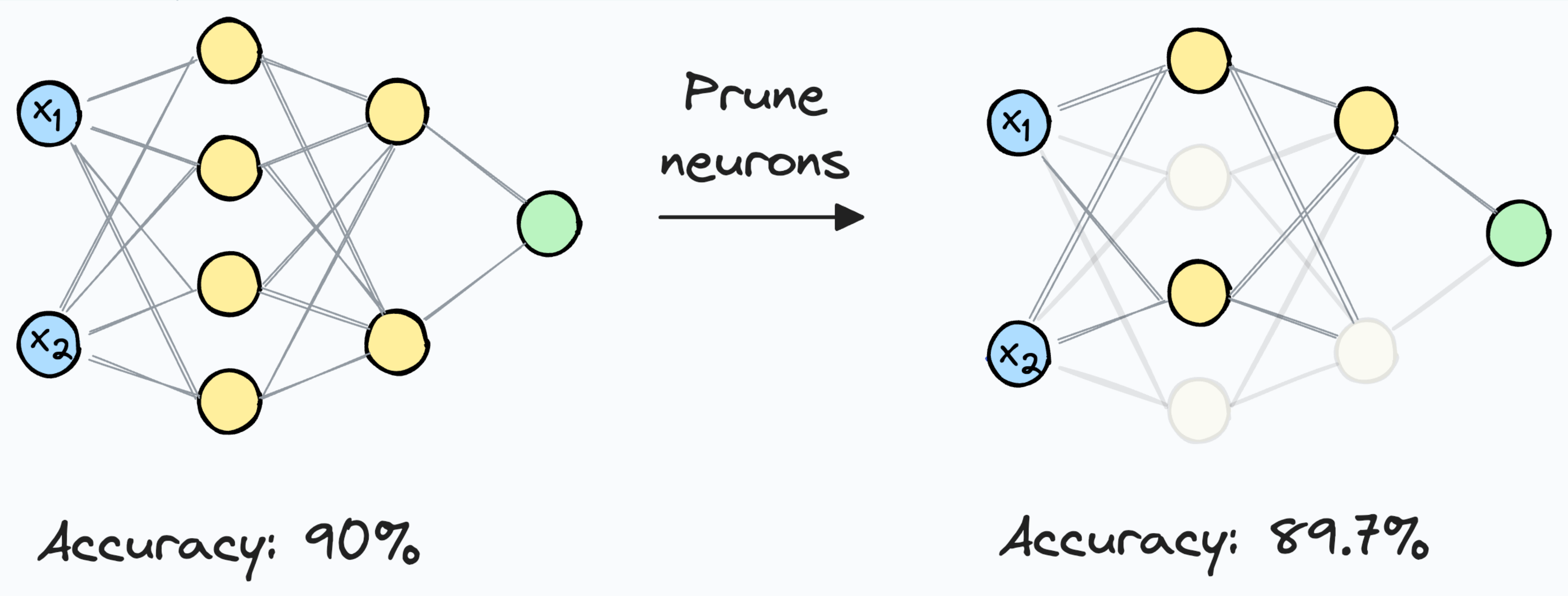
1x(36 RU) Hidden layers (HL), 36 in total

32 Features Maps (FM) per conv. layer

MPAI.
community

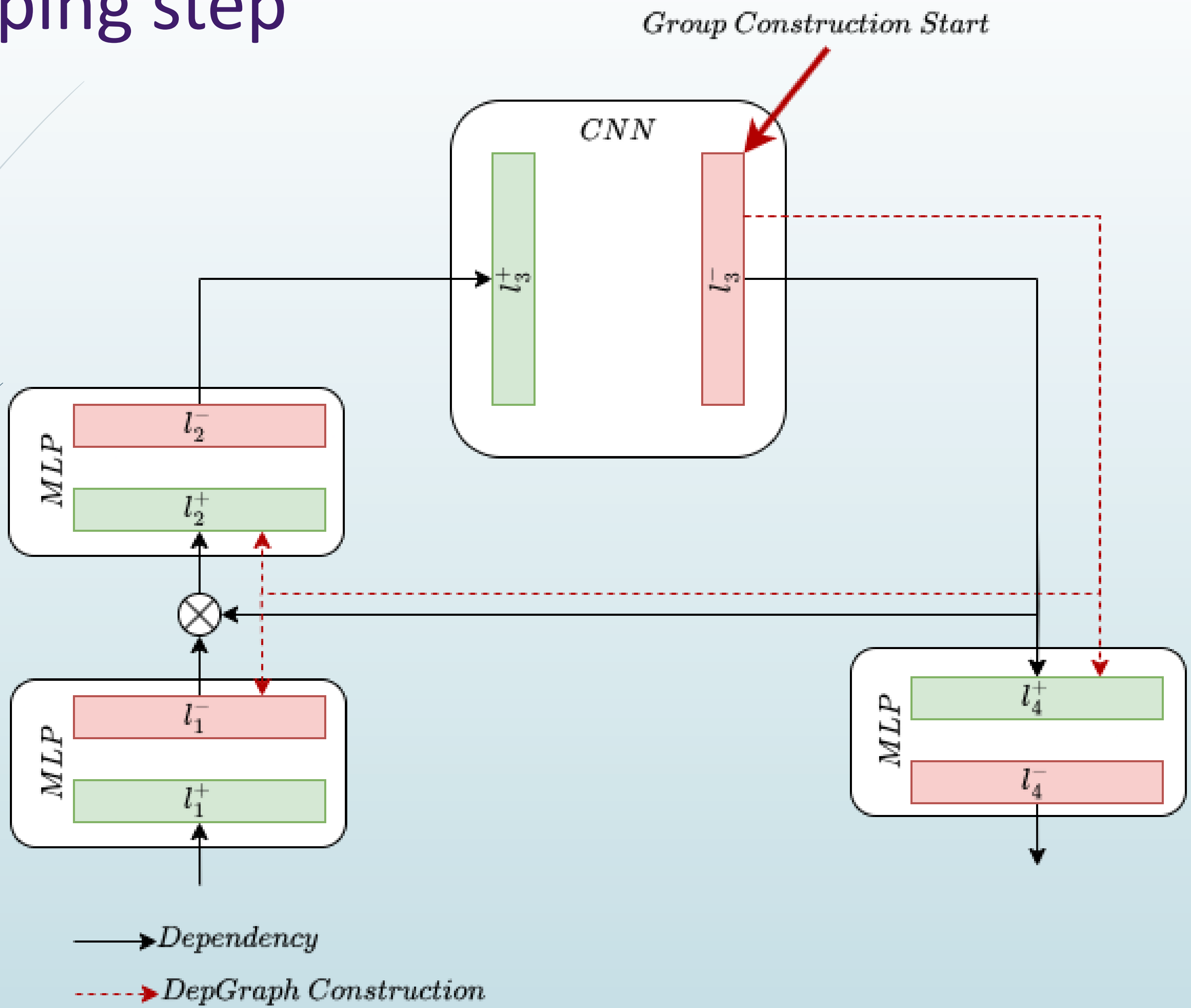
# Further Complexity Reduction (Pruning)

# General concept



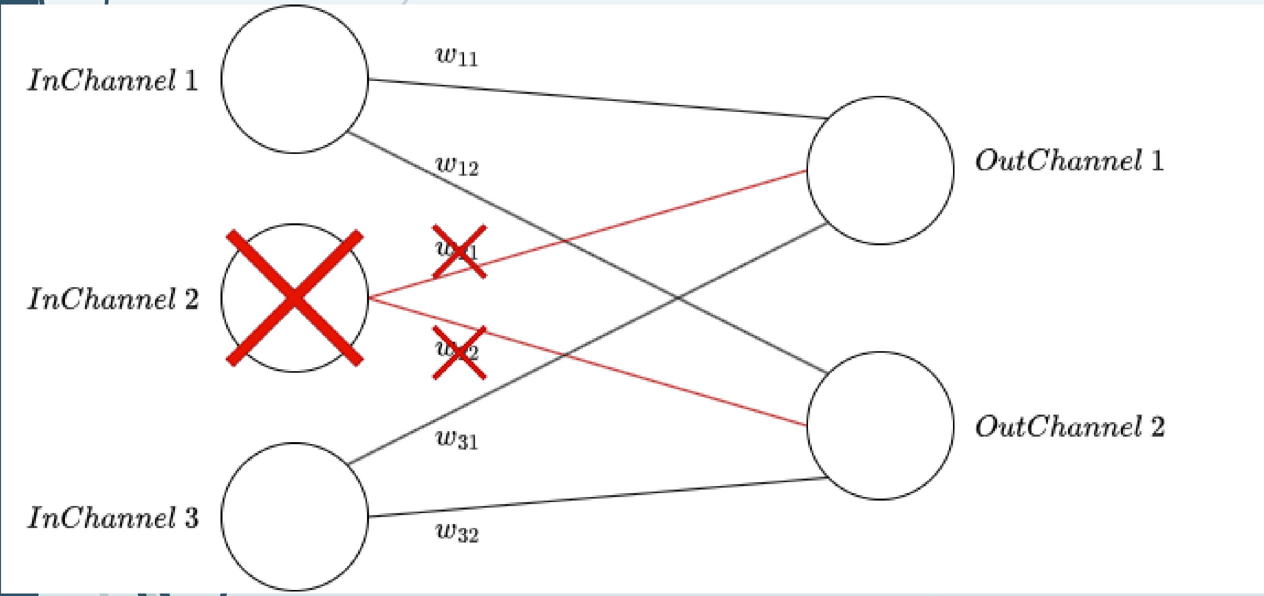

After: https://blog.dailydoseofds.com/p/activation-pruning-reduce-neural

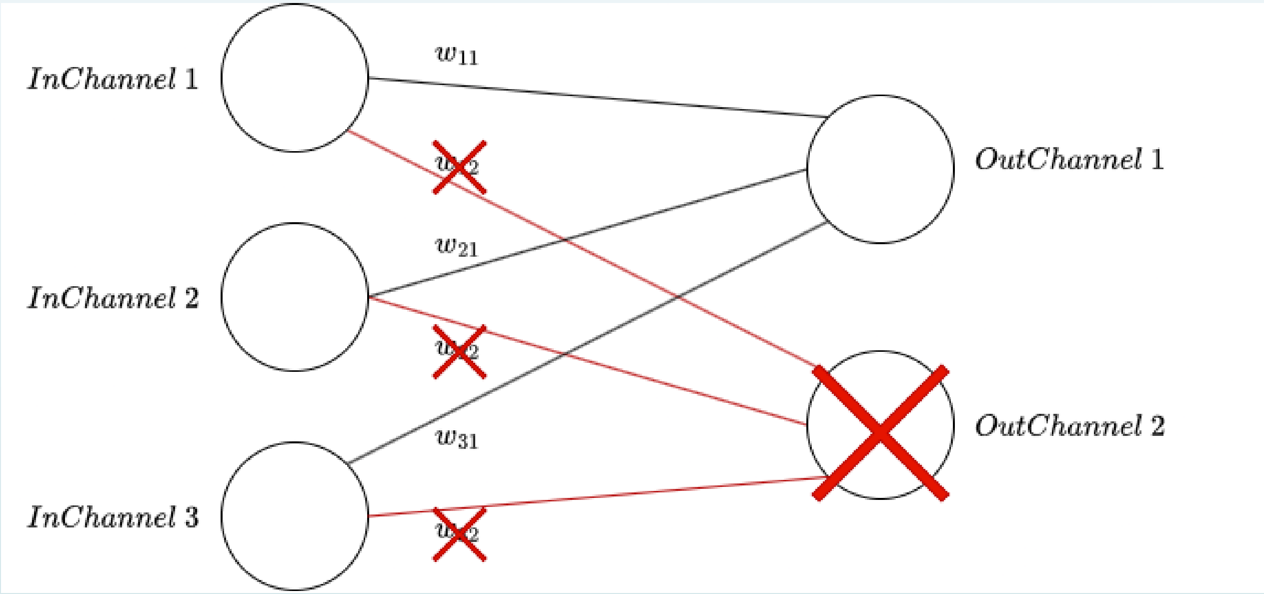
# Grouping step



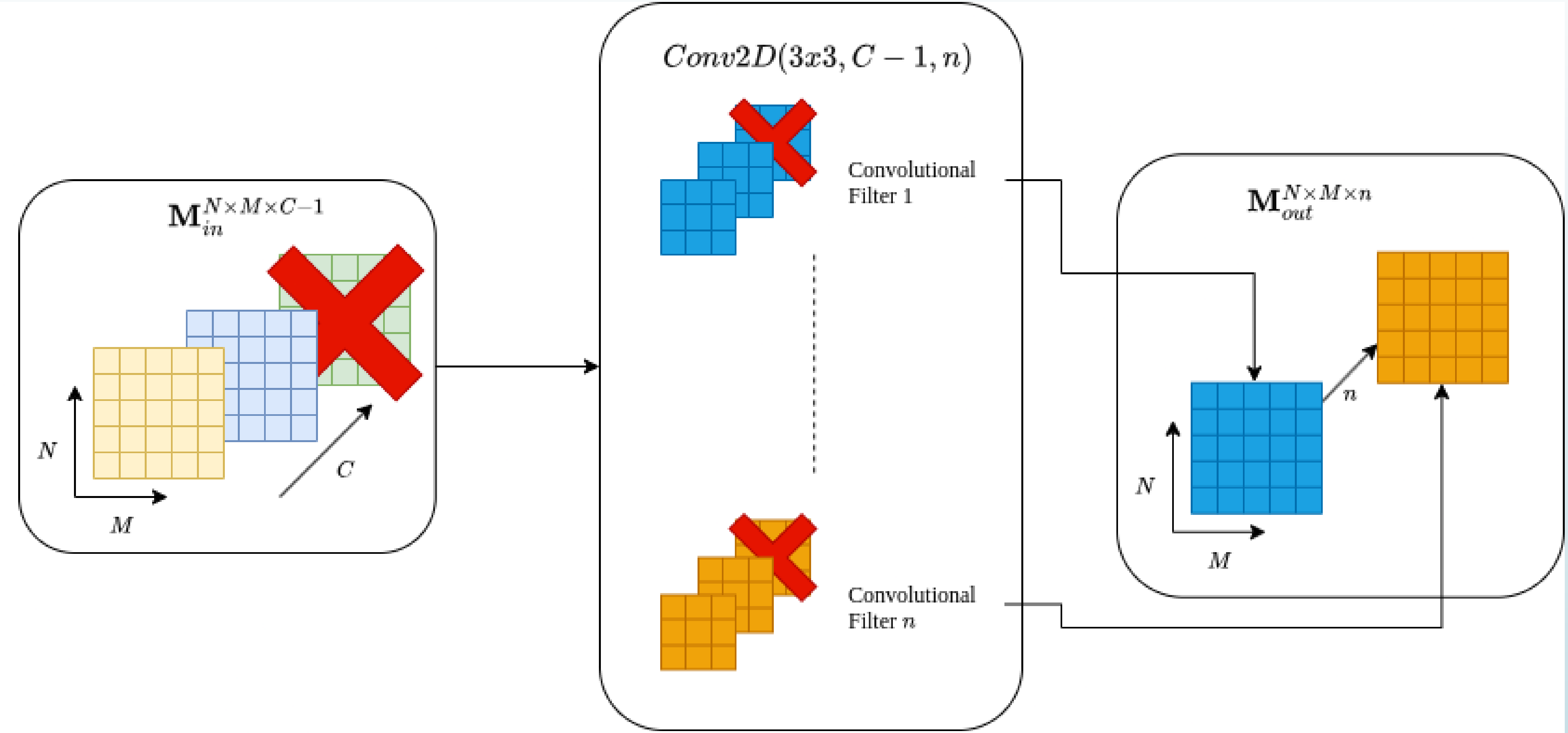
# Pruning Layers

# Fully connected Layer

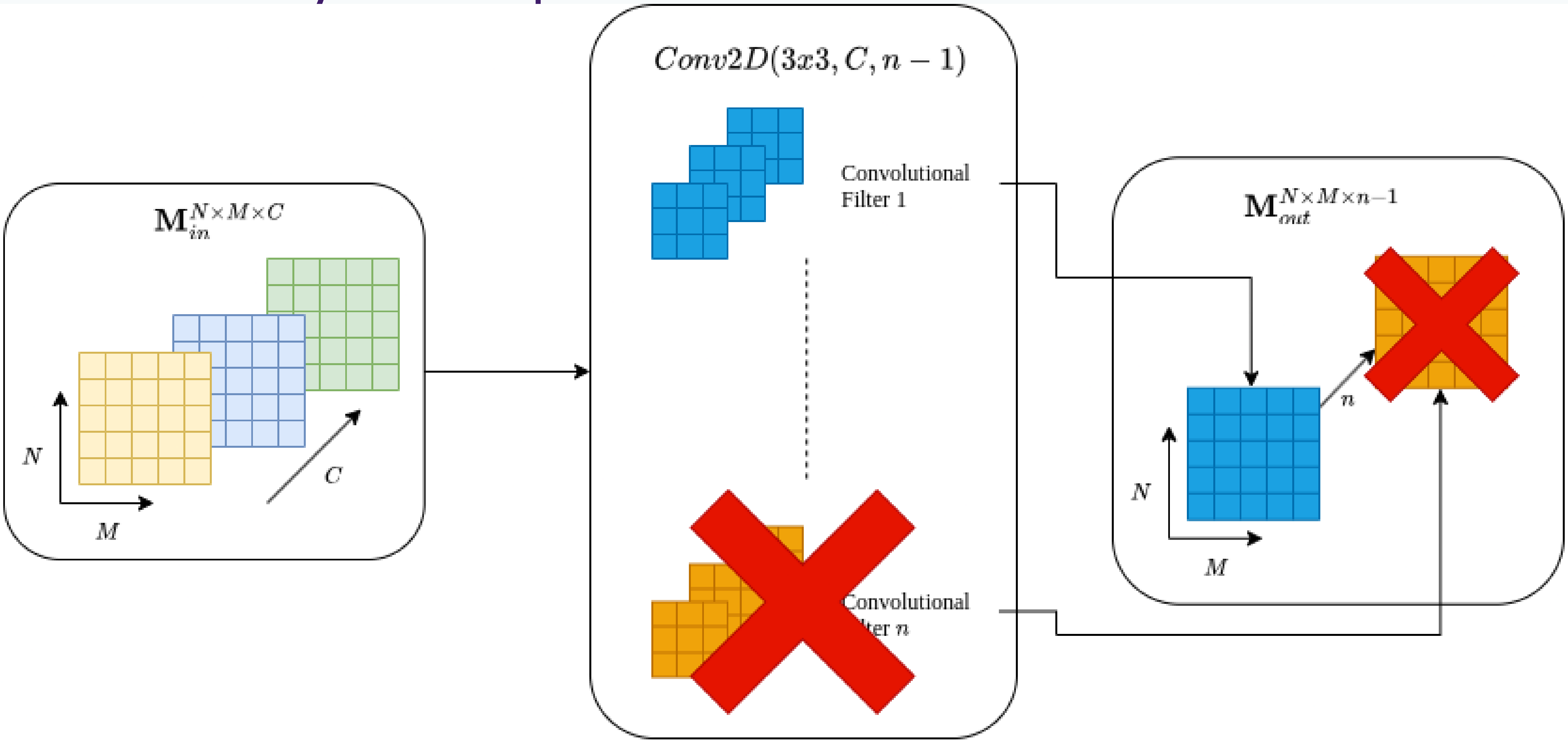


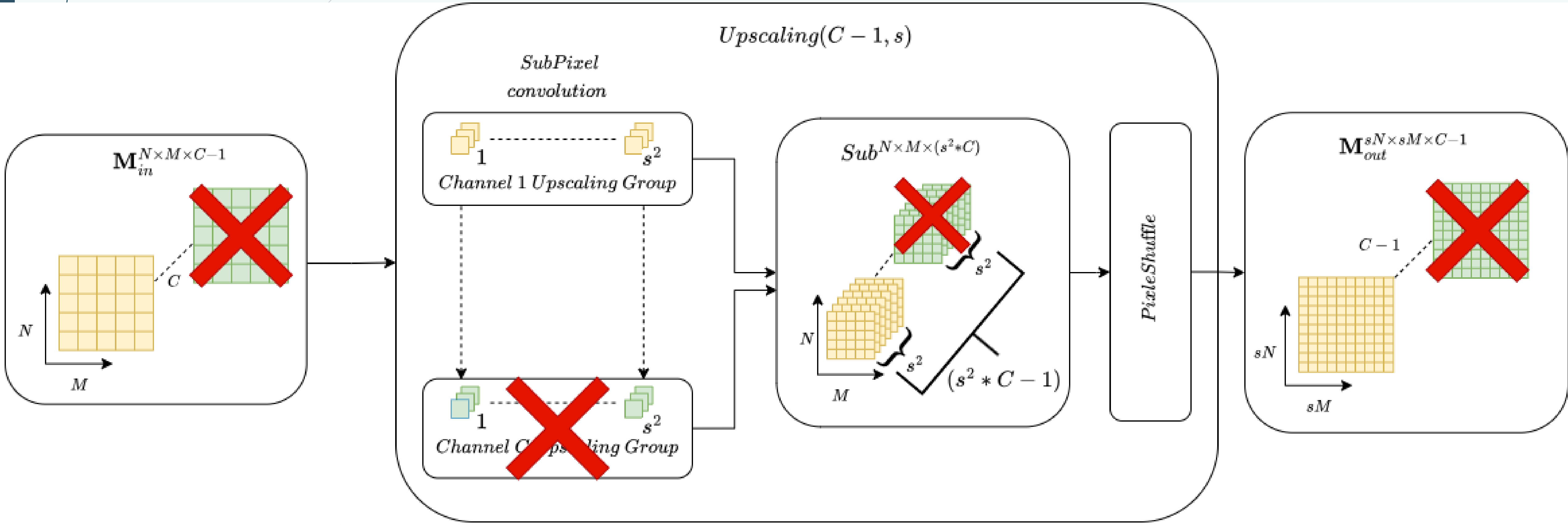
# CNN layer - Input



# CNN layer - Output



# Up-sampling Layer





community

# Performance of the EVC-UFV Up-sampling Filter

# Test Conditions

| | |
|---|---|
| **Standard sequences** | CatRobot, FoodMarket4, ParkRunning3. |
| **Encoding technology** | HEVC, VVC and AVC. |
| **Colous space** | YCbCr with 4:2:0 sub-sampling. |
| **Encoding settings** | Random Access and Low Delay at QPs 22, 27, 32, 37, 42, 47. |
| **Up-sampling** | SD to HD - HD to 4K. |
| **Metrics** | BD-Rate, BD-PSNR and BD-VMAF. |
| **Bits/sample** | 8 and 10 bit-depth per component. |

# Quality Evaluation (BD-rate)

LD = Low Delay
RA = Random Access

| Sequence name | HEVC (LD) | VVC (LD) | HEVC (RA) | VVC (RA) |
|---|---|---|---|---|
| SD to HD (using own trained filter) | 12.2% | 13.8% | 17.3% | 22.5% |
| HD to UHD (using own trained filter) | 6.0% | 6.5% | 6.0% | 7.9% |
| SD to HD (using HD to UHD filter) | 11.6% | 11.4% | 15.3% | 19.9% |

# Send your comments to

# [MPAI Secretariat]

# by 18 August 2025

# https://mpai.community/standards/mpai-evc/ufv/v1-0/

Join MPAI
Share the fun
Build the future

MPAI. community