Synamedia

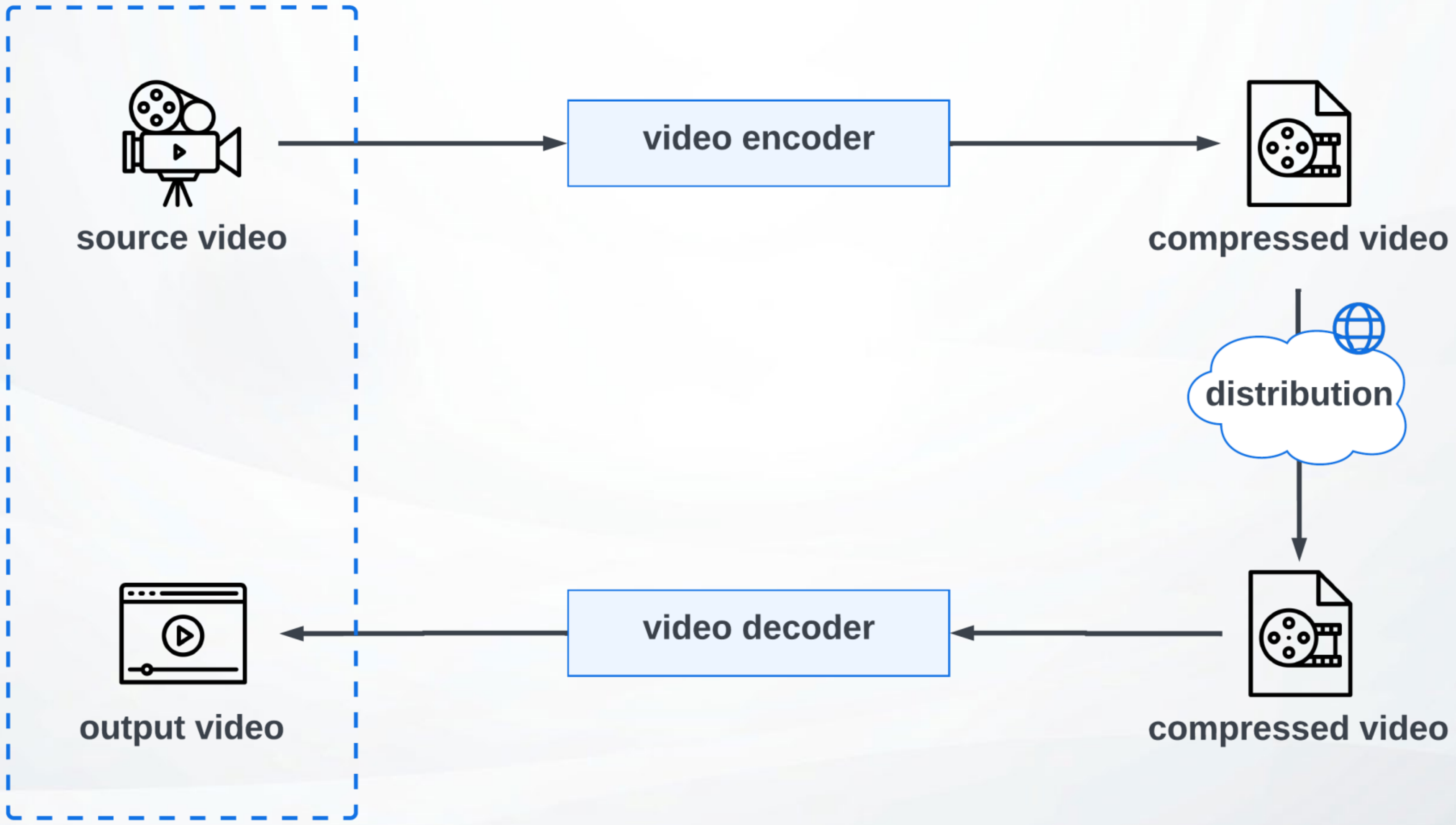
Combining Deep Learning and Feature Engineering for No-Reference Video Quality Assessment

Axel De Decker, Jan De Cock (Synamedia) & Glenn Van Wallendael (Ghent University)

Mile-High Video Conference 2024

February 12, 2024

Why we need No-Reference metrics



State-of-the-art

Feature-based metrics

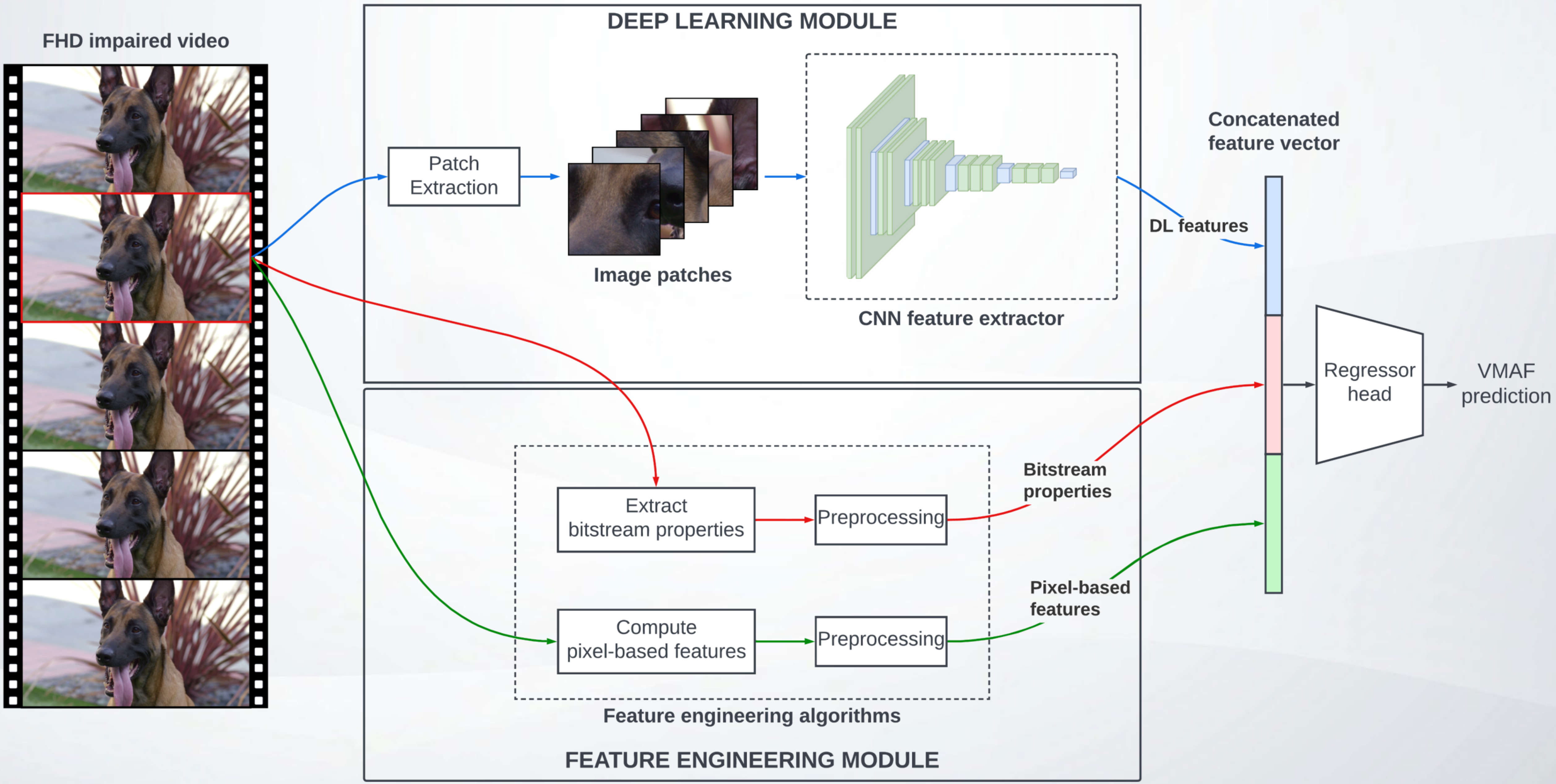
- Employ algorithmic solutions for feature extraction
- Advantages
  - Explainable AI
  - Low variance
- Disadvantages
  - Inaccurate

Deep Learning metrics

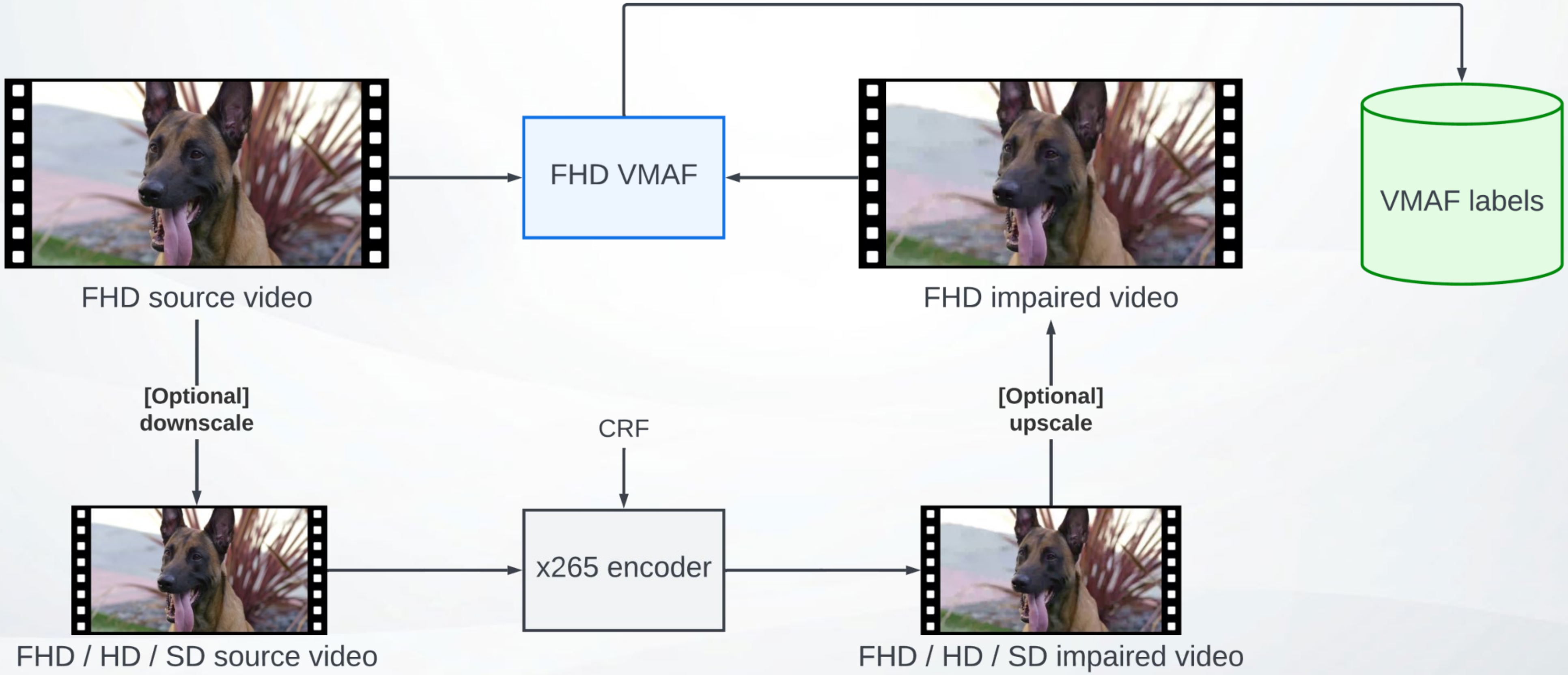
- Employ neural networks for feature extraction
- Advantages
  - Powerful predictive capabilities
- Disadvantages
  - Computationally expensive
  - Vulnerable to overfitting
  - Low interpretability

Hybrid metrics

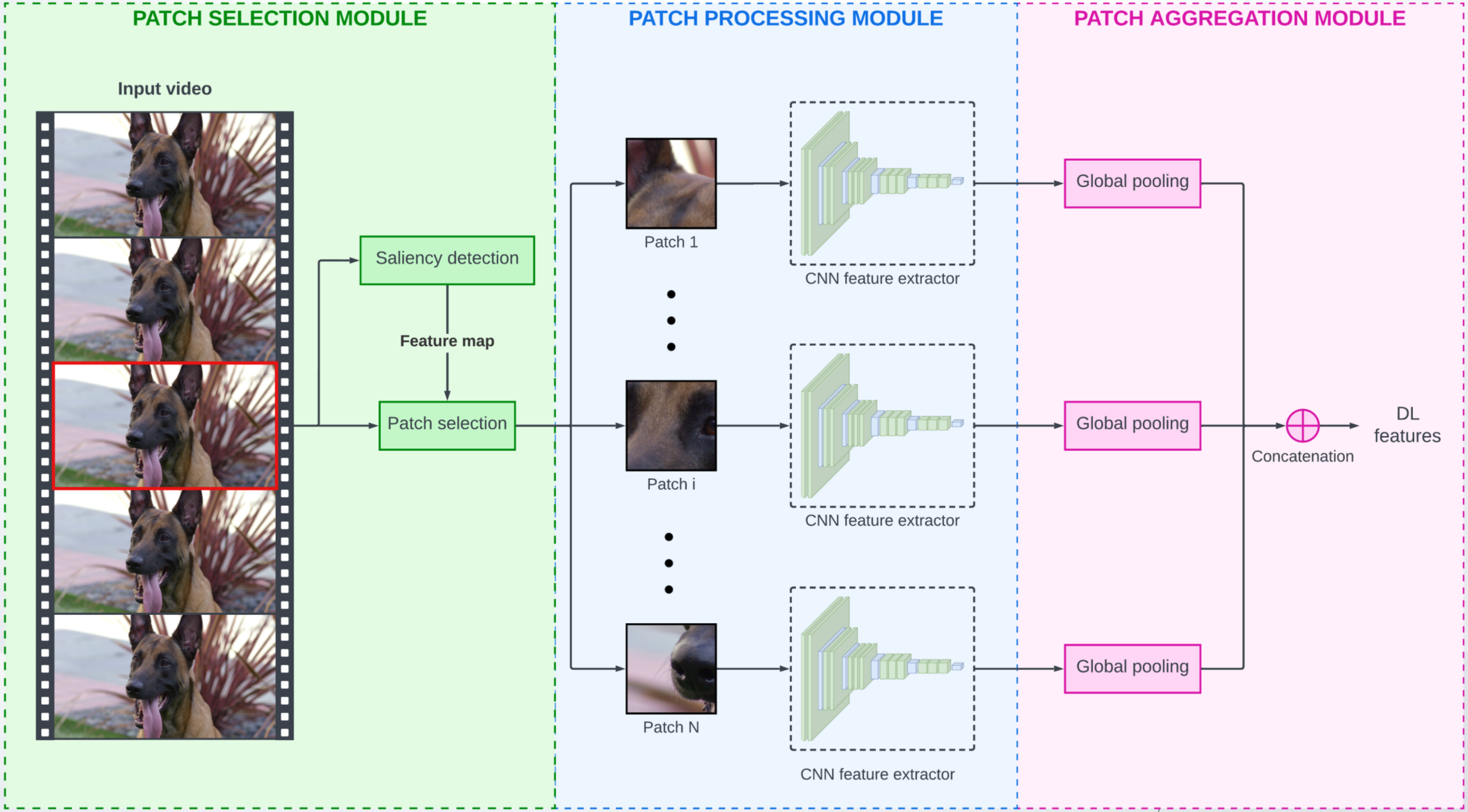
- Combine both methods for feature extraction
- Advantages
  - Benefits from strengths of both other types
- Disadvantages
  - Complex model architectures


FHD impaired video
DEEP LEARNING MODULE
Patch Extraction
Image patches
CNN feature extractor
DL features
Concatenated feature vector
Regressor head
VMAF prediction
Extract bitstream properties
Preprocessing
Bitstream properties
Compute pixel-based features
Preprocessing
Pixel-based features
Feature engineering algorithms
FEATURE ENGINEERING MODULE

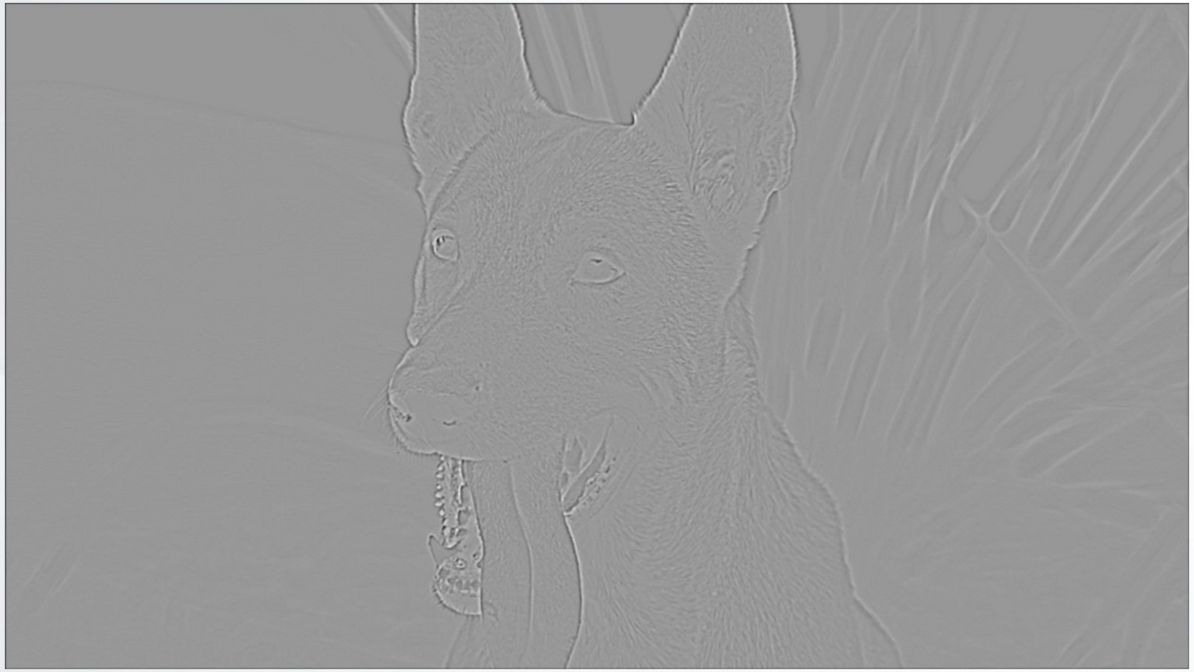
Data generation methodology

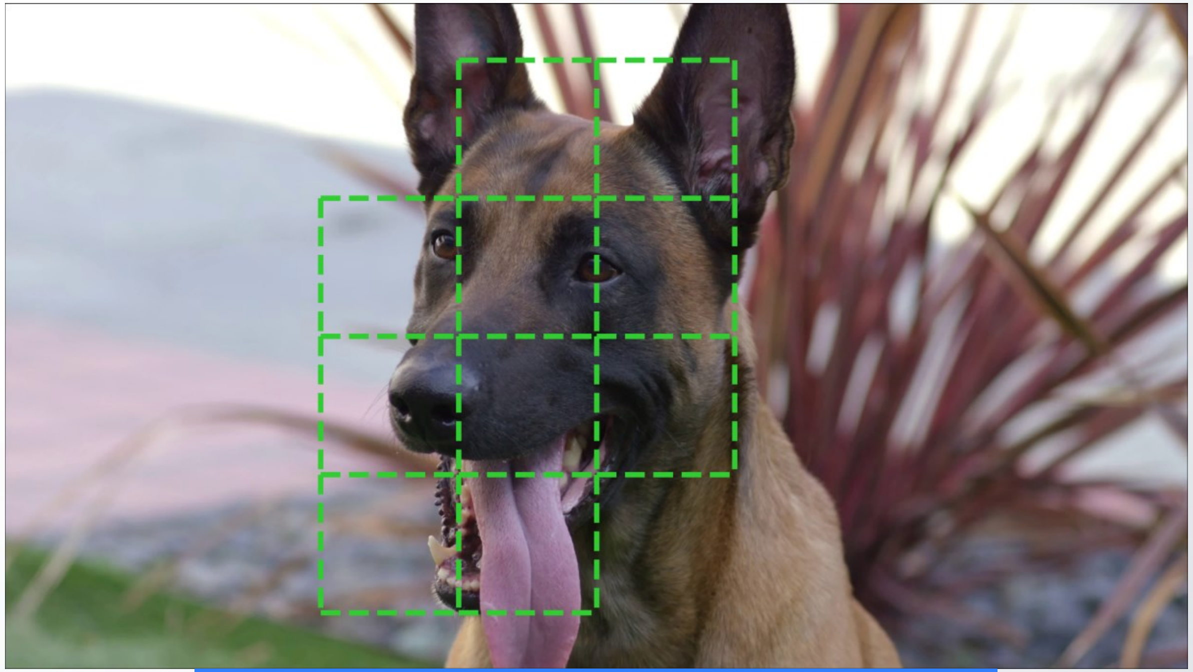


Extracting DL features from pixel data



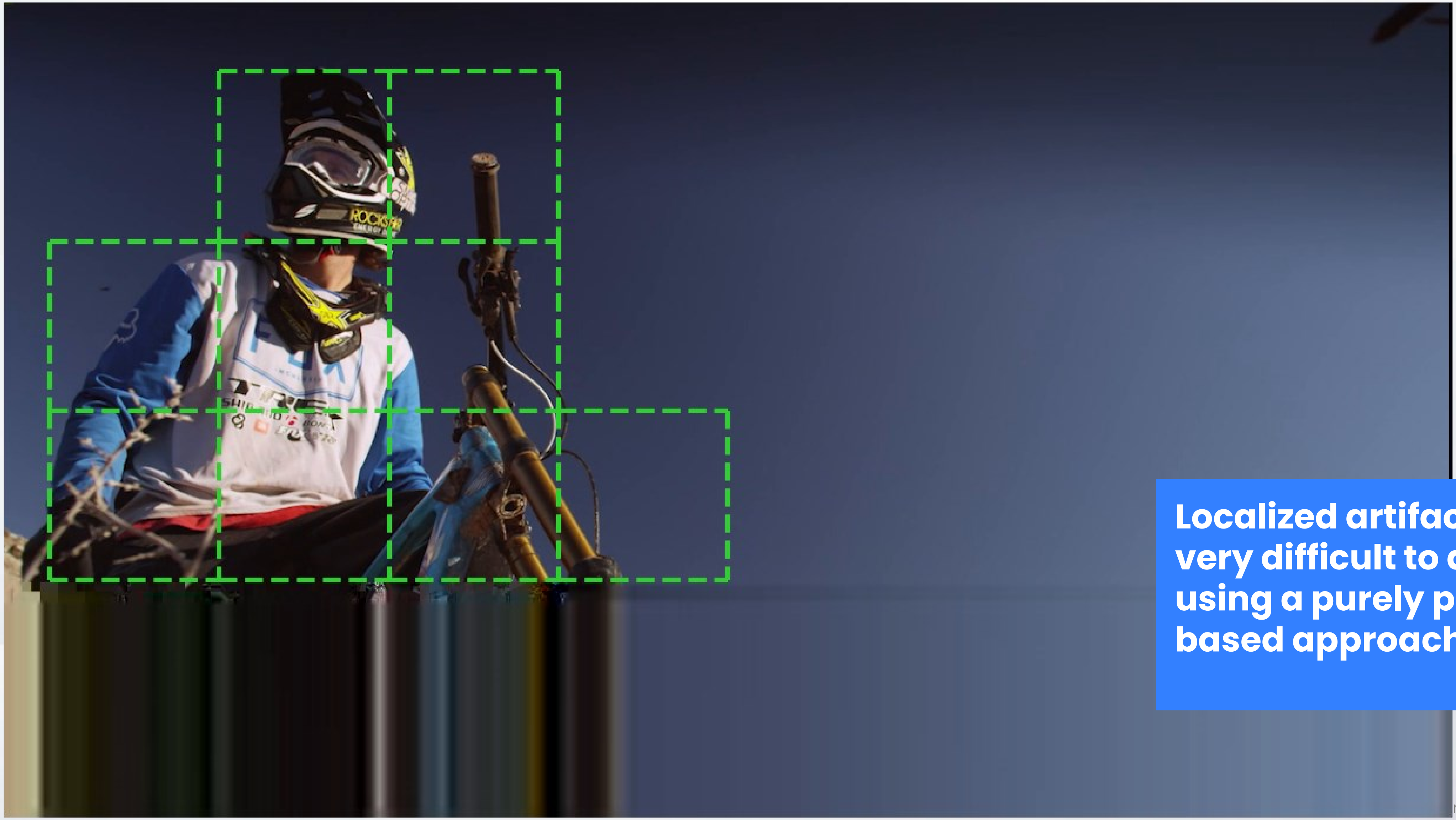
Patch selection





Predicted Region of Interest

Limitations of a patch-based approach



Localized artifacts are very difficult to detect using a purely patch-based approach!

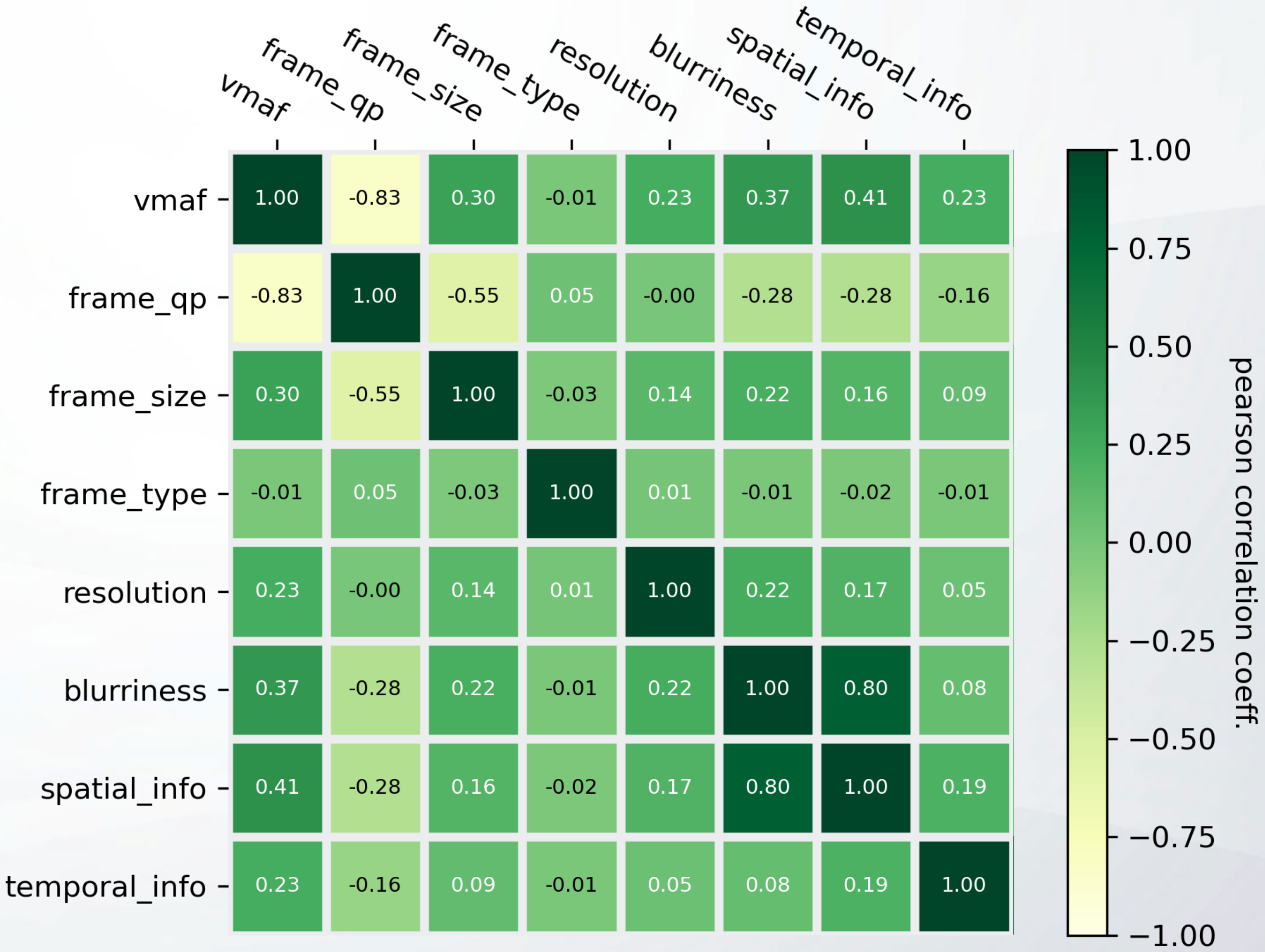
Handcrafted features

Bitstream properties

- Frame QP
- Frame size
- Frame type
- Resolution

Pixel-based features

- Blurriness
- Spatial information
- Temporal information



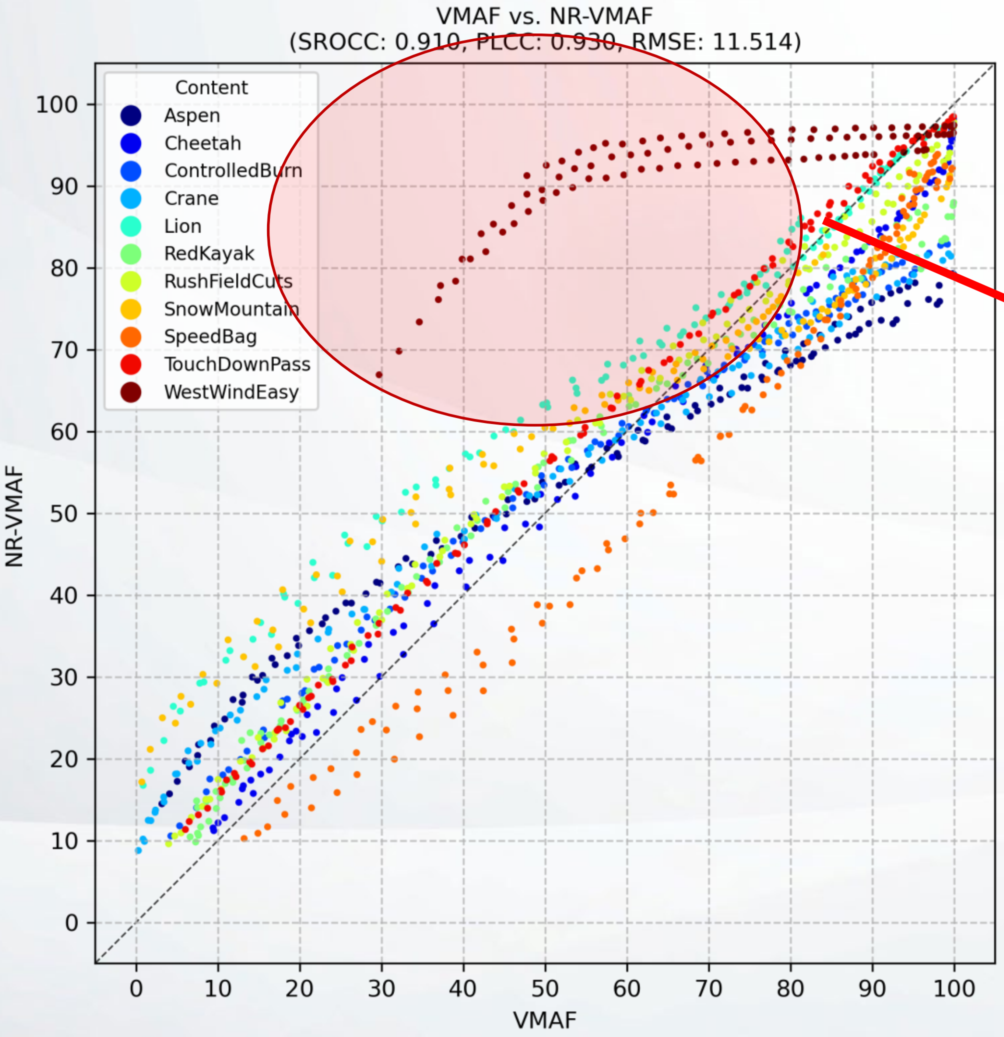
Reproducing VMAF

	Frame-level			Video-level		
	SROCC	**PLCC**	**MAE**	**SROCC**	**PLCC**	**MAE**
NR-VMAF	0.921	0.934	8.20	0.955	0.965	6.69
NR-PB-VMAF	0.972	0.970	5.88	0.986	0.985	4.54
Difference	+0.051	+0.037	-2.32	+0.031	+0.020	-2.15

Correlation and error measures between VMAF and NR-VMAF/NR-PB-VMAF averaged across different test datasets

Reproducing VMAF

DL model (NR-VMAF)

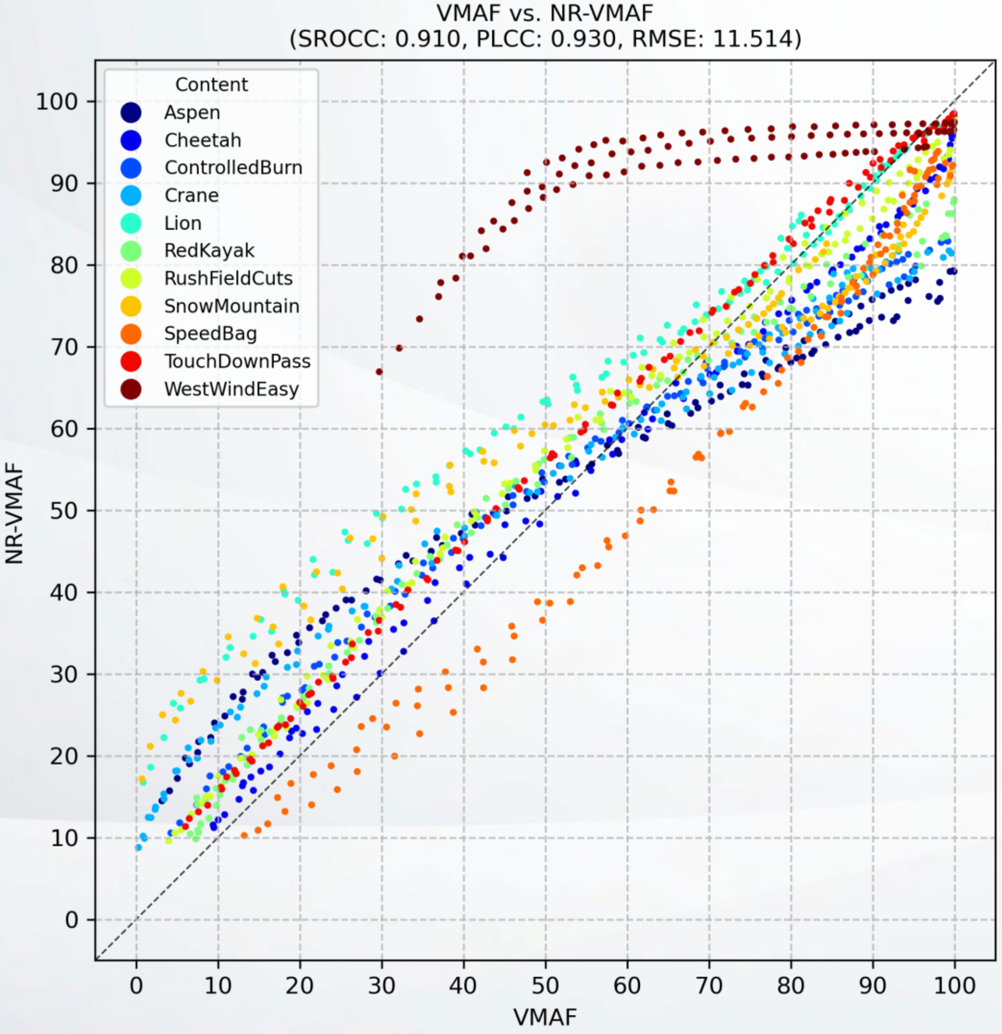


Severe overestimations of the actual VMAF scores, but why?

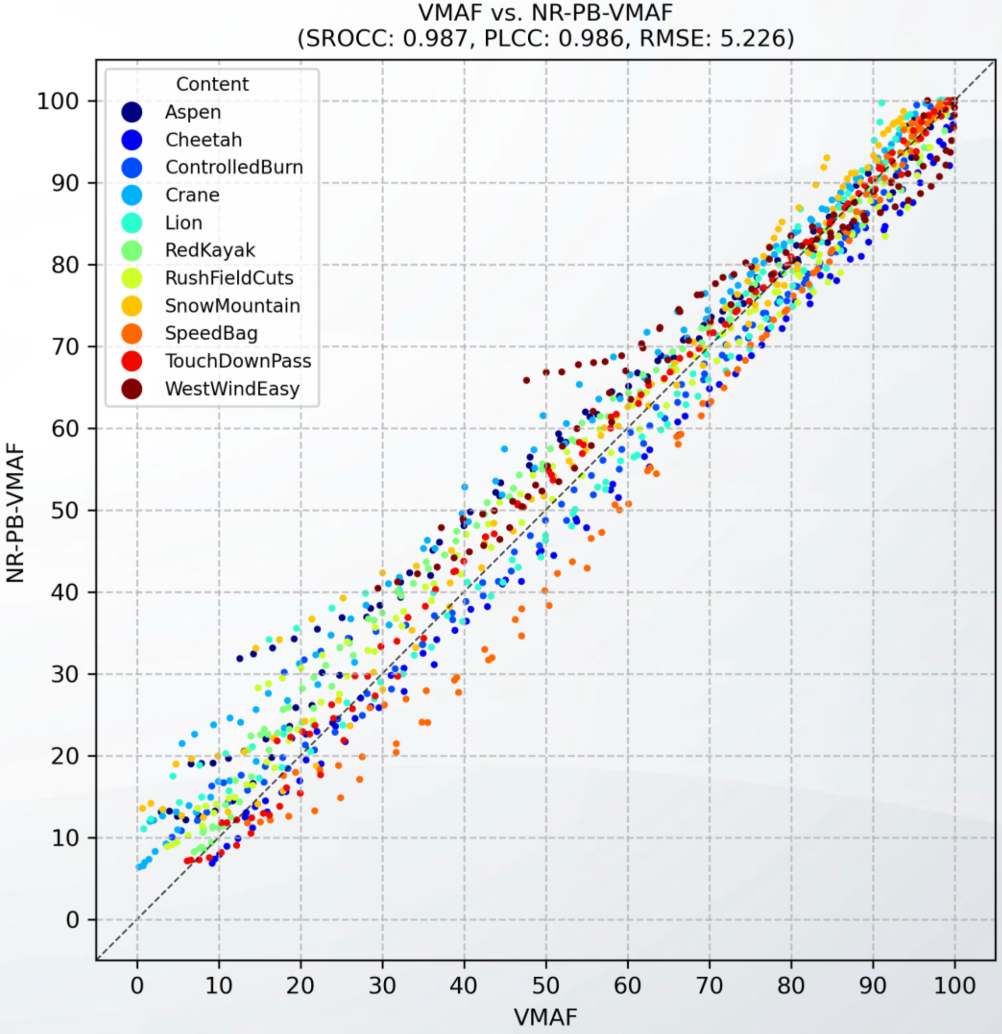
Limitations of a patch-based approach II

Reproducing VMAF

DL model (NR-VMAF)



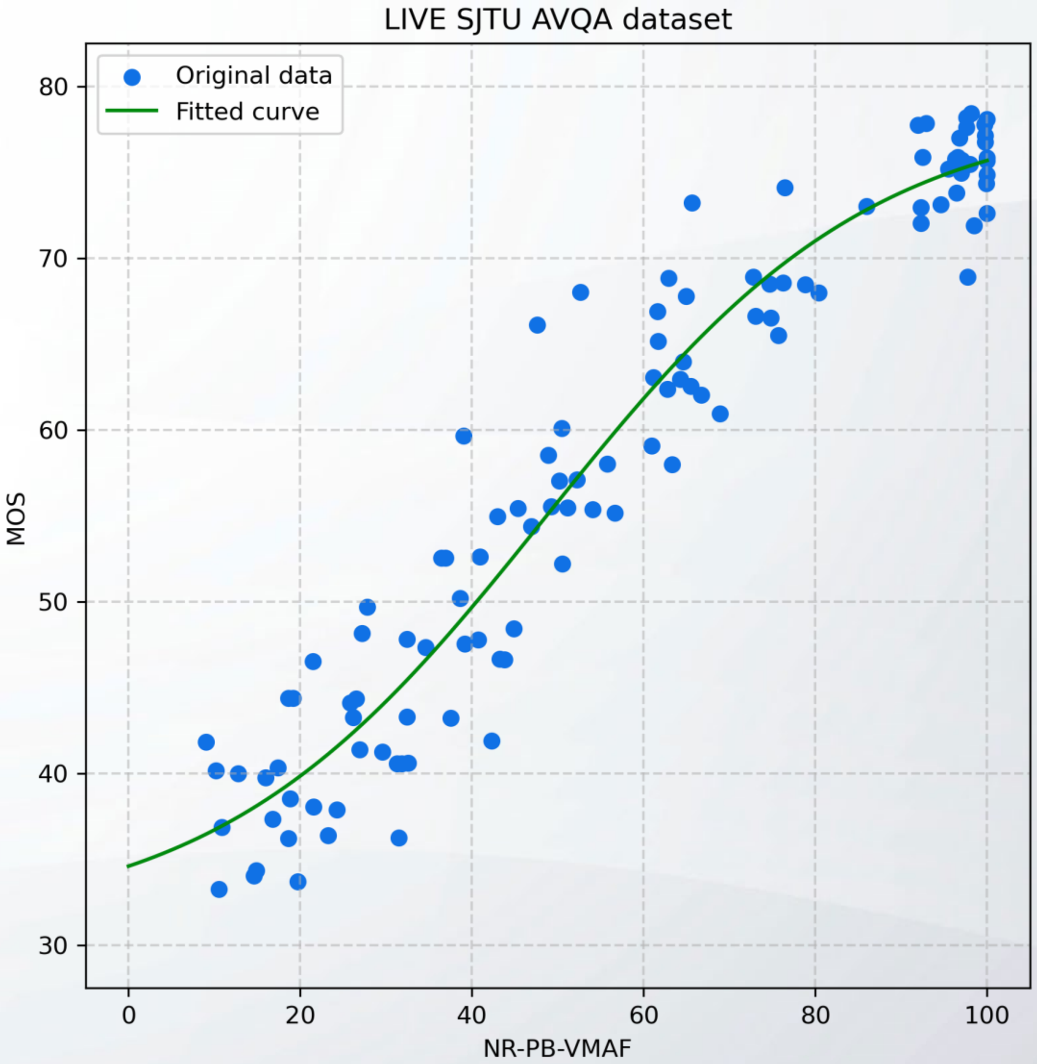
Hybrid model (NR-PB-VMAF)



Predicting human opinions

Cat.	Metric	SROCC	PLCC	RMSE
FR	PSNR	0.816	0.811	8.927
	SSIM	0.933	0.898	6.259
	MS-SSIM	0.914	0.885	6.639
	VMAF	0.946	0.959	4.028
NR	NIQE	0.334	0.244	13.771
	BRISQUE	0.631	0.579	11.592
	VIIDEO	0.068	0.127	14.080
	NR-VMAF	0.950	0.955	4.209
	NR-PB-VMAF	0.950	0.963	3.825

Correlation of various metrics with MOS on the LIVE SJTU AVQA dataset



Summary

- Deep learning architecture extracts informative features from patches

- Additional handcrafted features provide frame-level context
  - Advantages
    - Reduces outliers
    - Improves accuracy & robustness
  - Disadvantages
    - Makes the model more codec-specific
    - Cannot be (effectively) used on raw or transcoded videos

- Viable approach to No-Reference quality assessment <u>for certain use cases</u>

Synamedia

Q&A

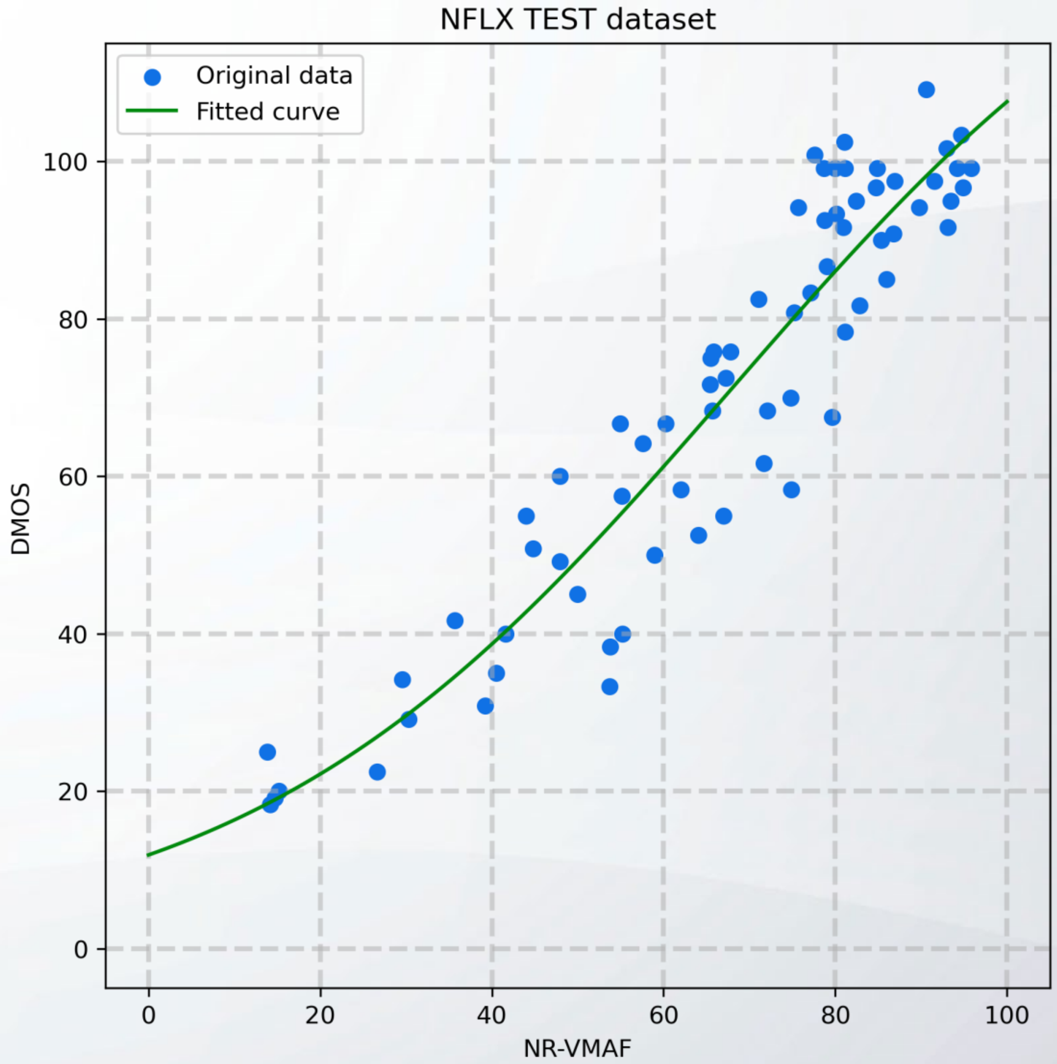
Synamedia

Extra slides

Predicting human opinions II

Cat.	Metric	SROCC	PLCC	RMSE
FR	PSNR	0.660	0.701	18.234
	SSIM	0.766	0.748	17.063
	MS-SSIM	0.738	0.754	16.875
	VMAF	0.911	0.935	9.071
NR	NIQE	0.702	0.671	19.042
	BRISQUE	0.785	0.832	14.232
	VIIDEO	0.548	0.555	21.353
	NR-VMAF	0.910	0.937	8.969

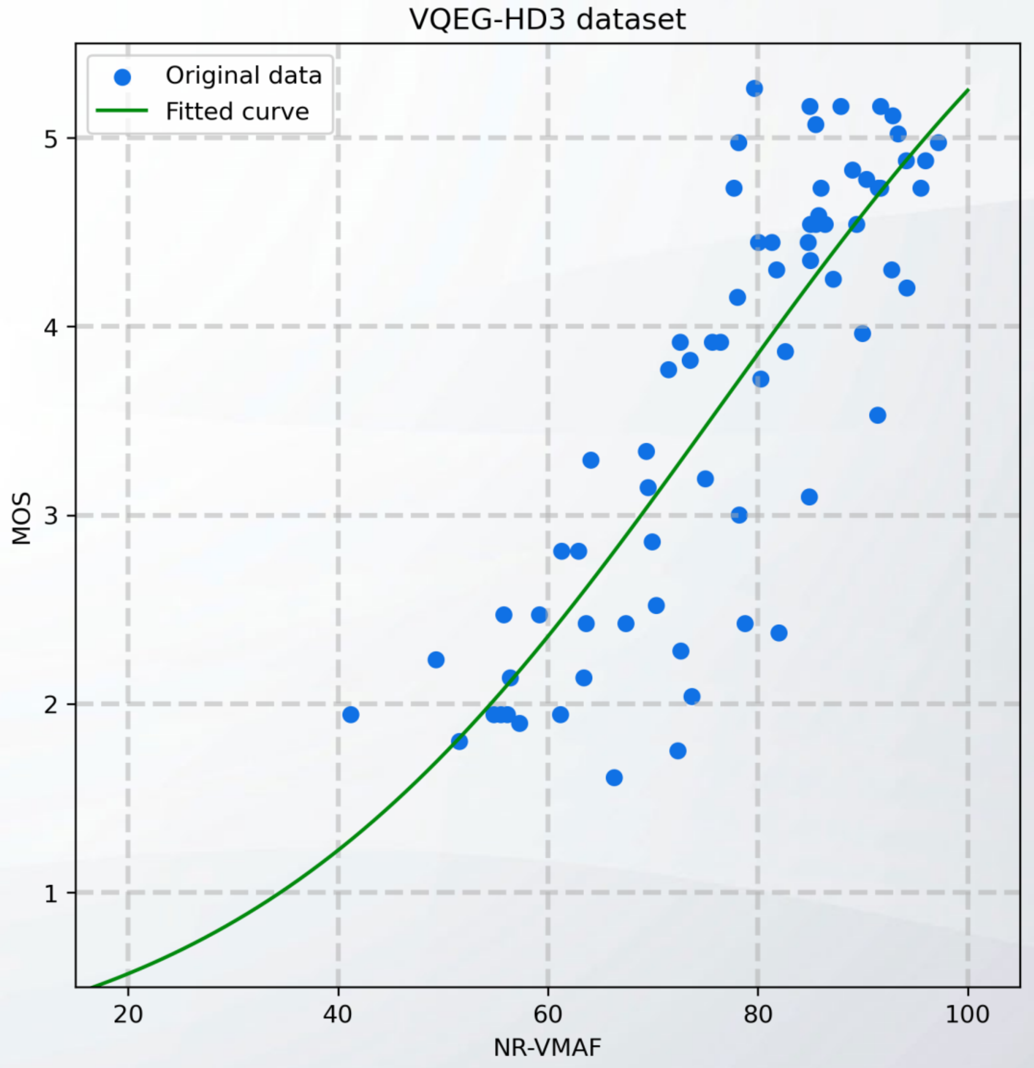
Correlation of various metrics with DMOS on the NFLX TEST dataset



Predicting human opinions III

Cat.	Metric	SROCC	PLCC	RMSE
FR	PSNR	0.784	0.787	0.697
	SSIM	0.906	0.880	0.539
	MS-SSIM	0.897	0.873	0.552
	VMAF	0.921	0.934	0.404
NR	NIQE	0.264	0.068	1.139
	BRISQUE	0.535	0.523	0.973
	VIIDEO	0.092	0.087	1.138
	NR-VMAF	0.813	0.834	0.630

Correlation of various metrics with MOS on the VQEG-HD3 dataset



Future directions

1. Increase the computational efficiency
   - Experiment with lightweight CNN architectures
   - Drop redundant features
   - Advantages
     - Faster inference
     - More energy-efficient
     - Reduced training times

2. Increase the applicability of our metric
   - New distortion types
   - New devices
   - Varying video characteristics

Synamedia