

# OBJECTIVE MEASURES ON THE ITS4S DATABASE

---

Enrico Masala

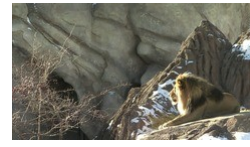
Politecnico di Torino, Italy

[enrico.masala@polito.it](mailto:enrico.masala@polito.it)

VQEG JEG-Hybrid session in Mountain View, Nov 2018

# JEG-Hybrid Context

- Objectively-annotated Large Scale Database
  - 59,520 HEVC-encoded video sequences (1,920 HRCs)
    - 10 sources, 250 frames each, 25 fps
    - 3 resolutions: 1920x1080, 1280x720, 960x544  
(details in references, already presented in previous meetings)
    - PSNR, SSIM, MS-SSIM, VIF, VQM, VMAF (0.6.0, 0.6.1), PVQM
    - Distortion due to encoding
    - Distortion due to encoding + data (packet) loss (~500,000 samples)



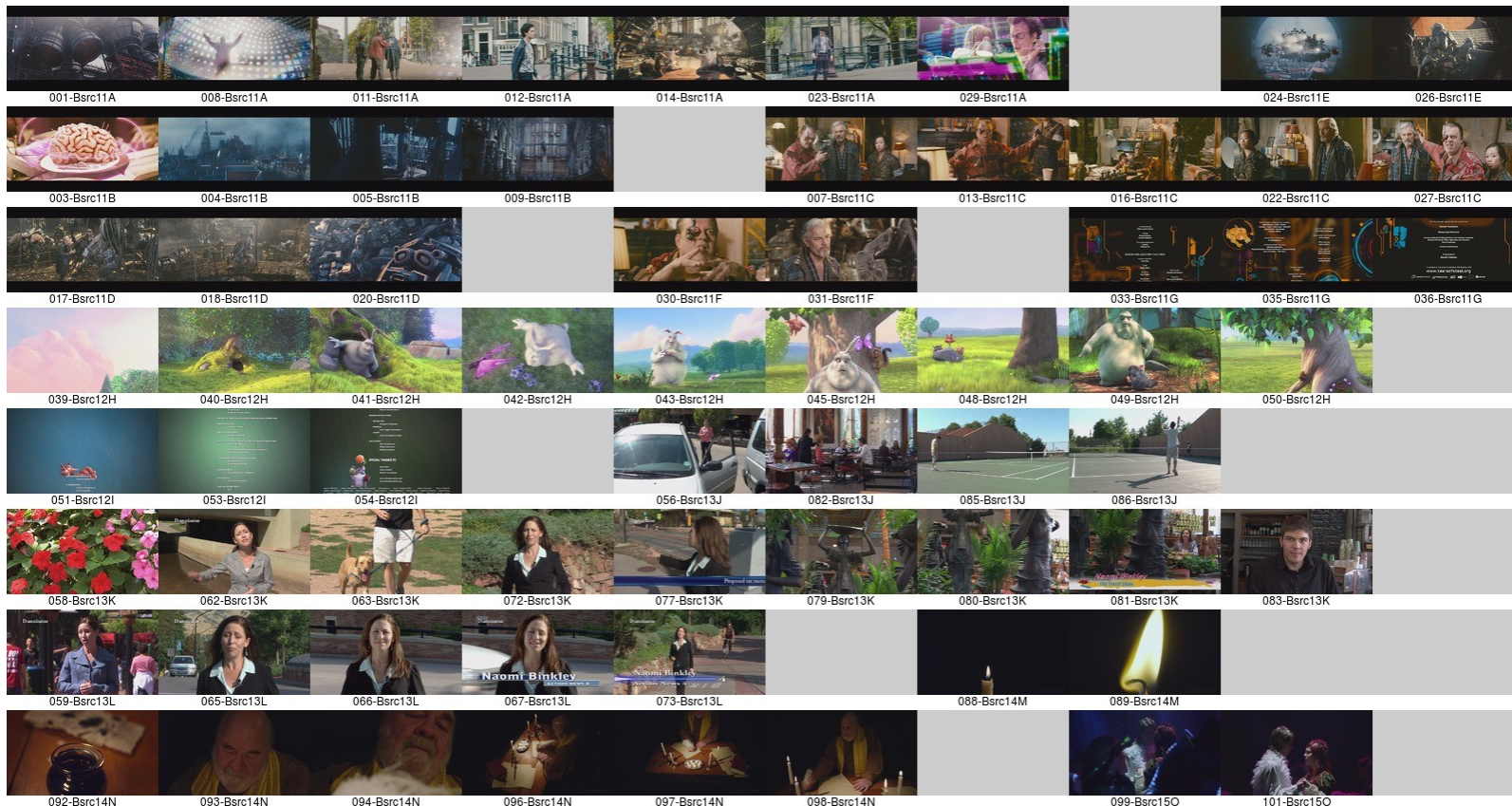
# Subjective Annotation?

- ITS4S database
  - 4-second length sequences, 1280x720, 24 fps
  - 813 sequences (from 35 video footages)
    - 8 themes + 1 miscellaneous (Broadcast, Everglades, MusicMexico, Nature, Ocean, PublicSafety, Sports, Training)
  - Original purpose: No-Reference Study
  - **Subjectively annotated** (ACR) by 30+ subjects
  - **Original footage is available for all sequences**
- Idea
  - Run objective quality measures as in the JEG-Hybrid large-scale DB, but having subjective annotation
  - Suitable for this purpose:
    - 518 sequences, 5 HRCs (compression artifacts due to different coding bitrate)
    - A minor share are deemed “bad quality” from the start
    - Note: PVS have to be temporally aligned

# Sample of Sequences

- Broadcast
- Chance (miscellaneous)
- Everglades
- Music&Mexico
- Nature
- Ocean
- PublicSafety
- Sports
- Training

# Broadcast



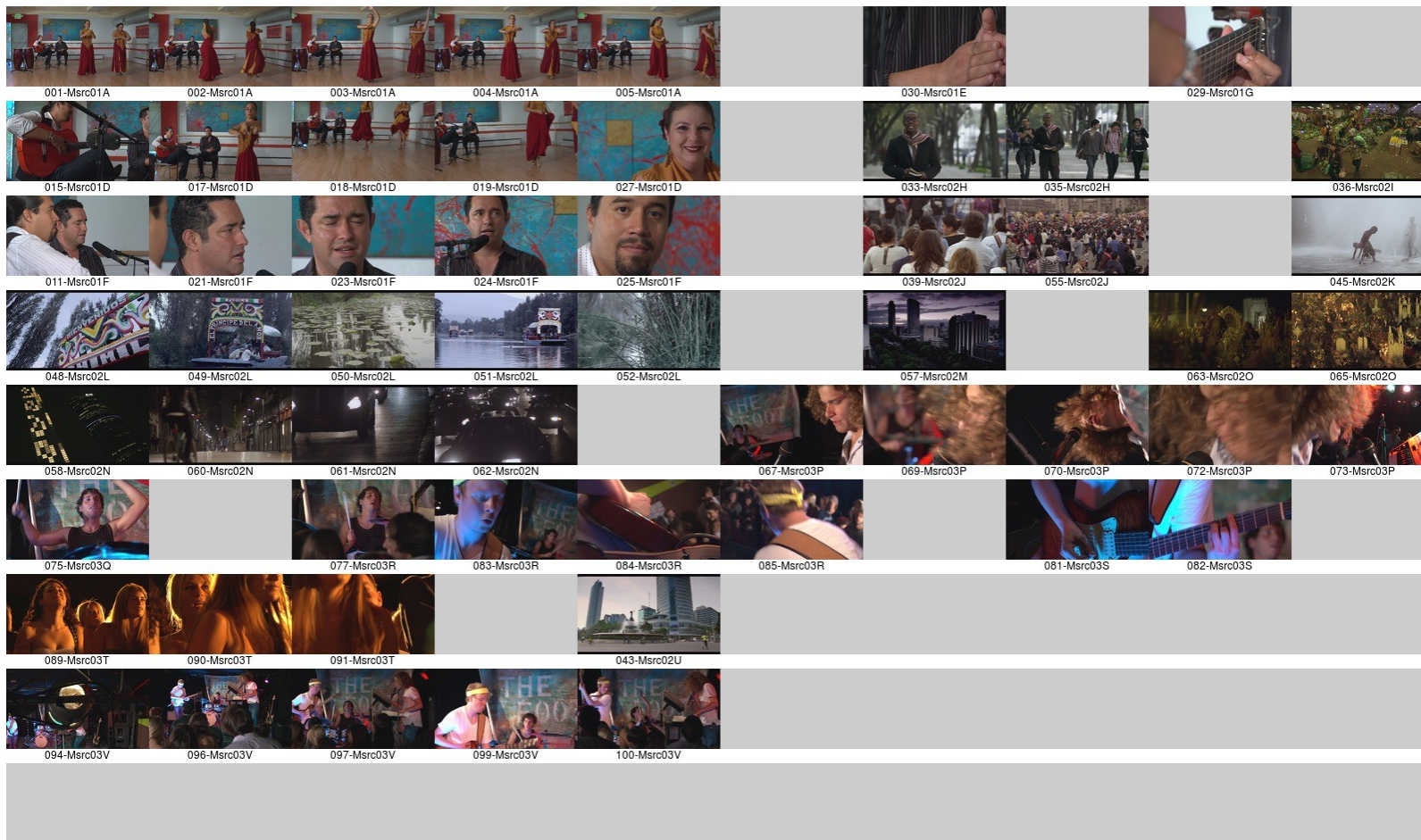
# Chance (Miscellaneous)



# Everglades



# Music & Mexico

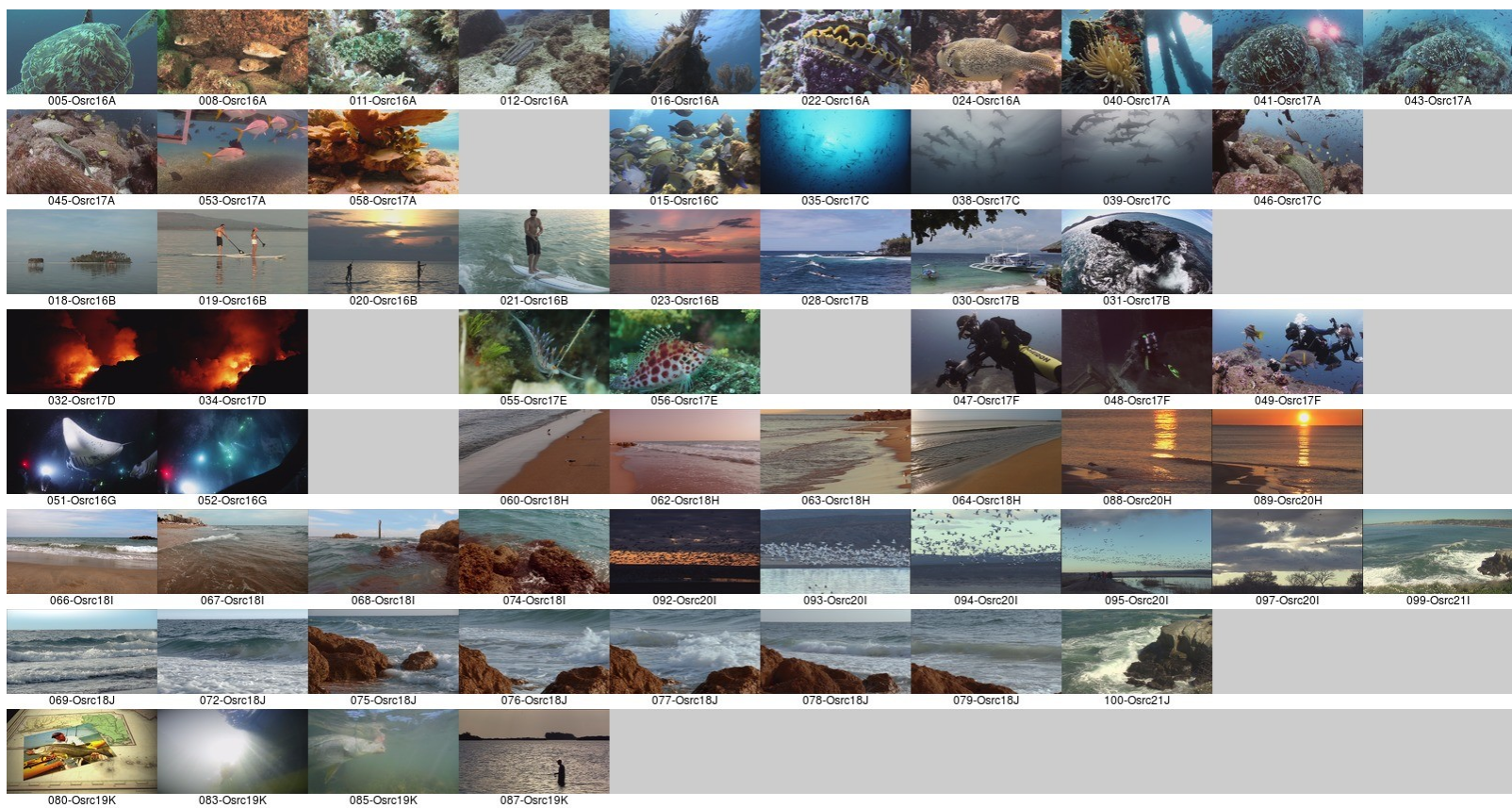




# Nature



# Ocean



# Public Safety



# Sports



# Training

001-Tsrc35A

003-Tsrc35A

005-Tsrc35A

006-Tsrc35A

007-Tsrc35A

008-Tsrc35B

009-Tsrc35B

010-Tsrc35B

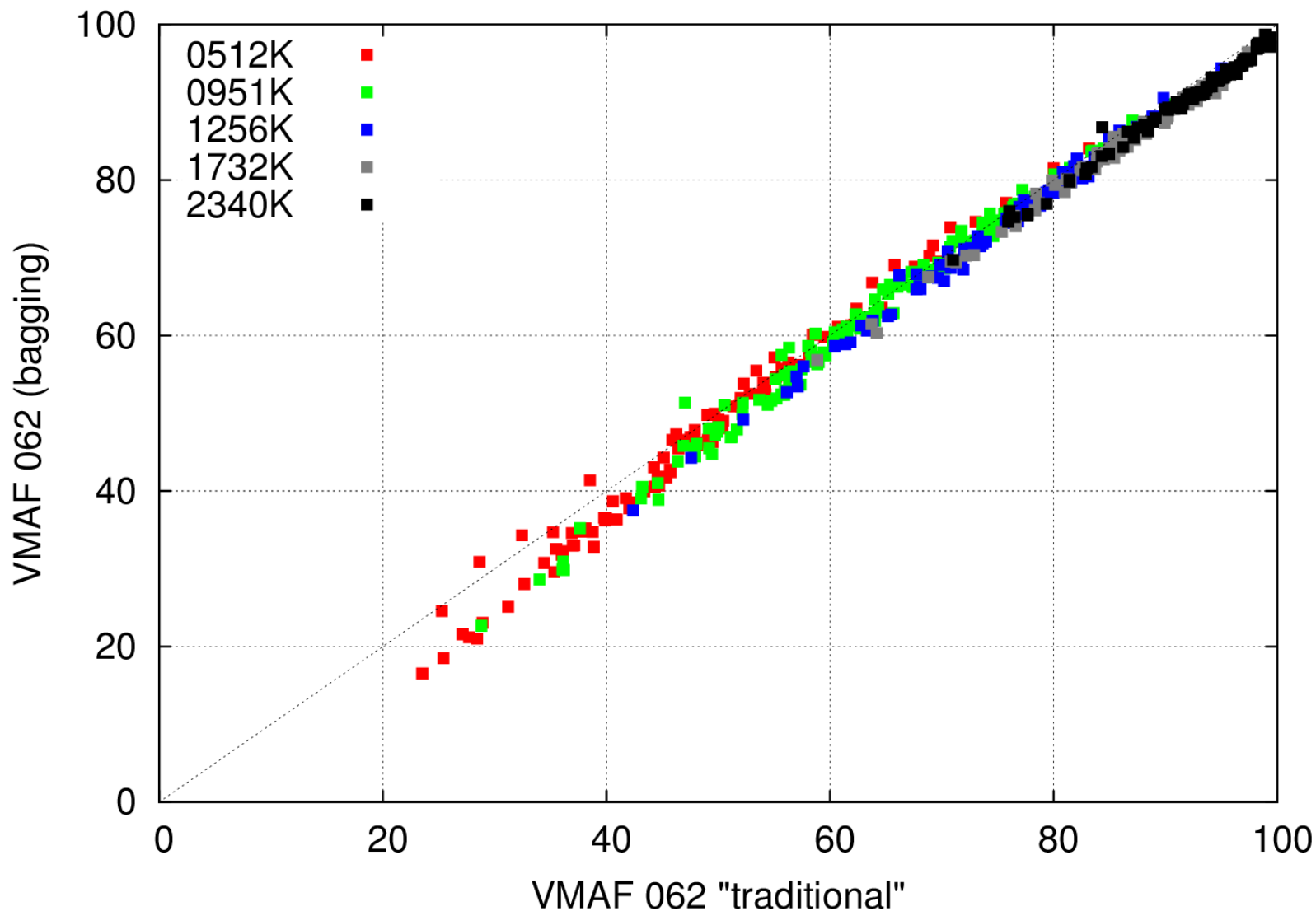
011-Tsrc35C

The image strip contains 11 small images. The first image (001-Tsrc35A) shows a rocket launch at night. The second (003-Tsrc35A) shows a bright light source in a dark environment. The third (005-Tsrc35A) shows a rocket on a launch pad. The fourth (006-Tsrc35A) shows a rocket launch with a large plume of smoke. The fifth (007-Tsrc35A) shows a rocket on a launch pad. The sixth (008-Tsrc35B) shows a rocket launch with a large plume of smoke. The seventh (009-Tsrc35B) shows a rocket launch with a large plume of smoke. The eighth (010-Tsrc35B) shows a rocket launch with a large plume of smoke. The ninth (011-Tsrc35C) shows a rocket launch with a large plume of smoke. Below the strip is a large grey rectangular area.

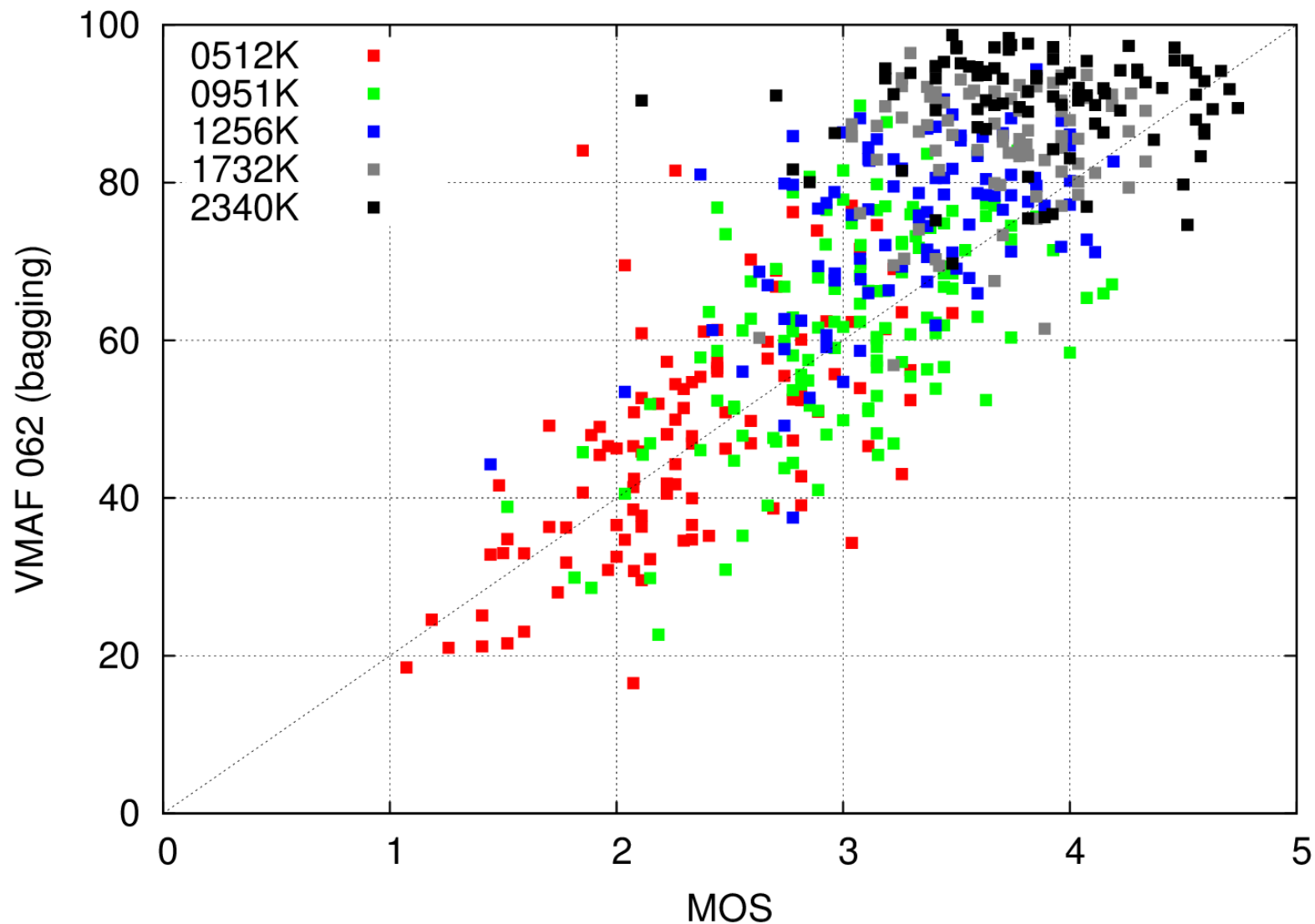
# Objective Measures

- PSNR, SSIM, MS-SSIM, VIF (from vqmt by EPFL)
- VMAF 0.6.0, 0.6.1
- VMAF 0.6.2 traditional and resulting from bootstrap aggregation (“bagging”), with 95% confidence intervals
- PSNR, SSIM, MS-SSIM (from vmaf by Netflix)
- VQM
  - In some cases, sequences have a few repeated frames to reach 4 sec length needed to run the executable-only vqm version
- PVQM
  
- Some results in the following

# VMAF 0.6.2 (bagging vs “traditional”)

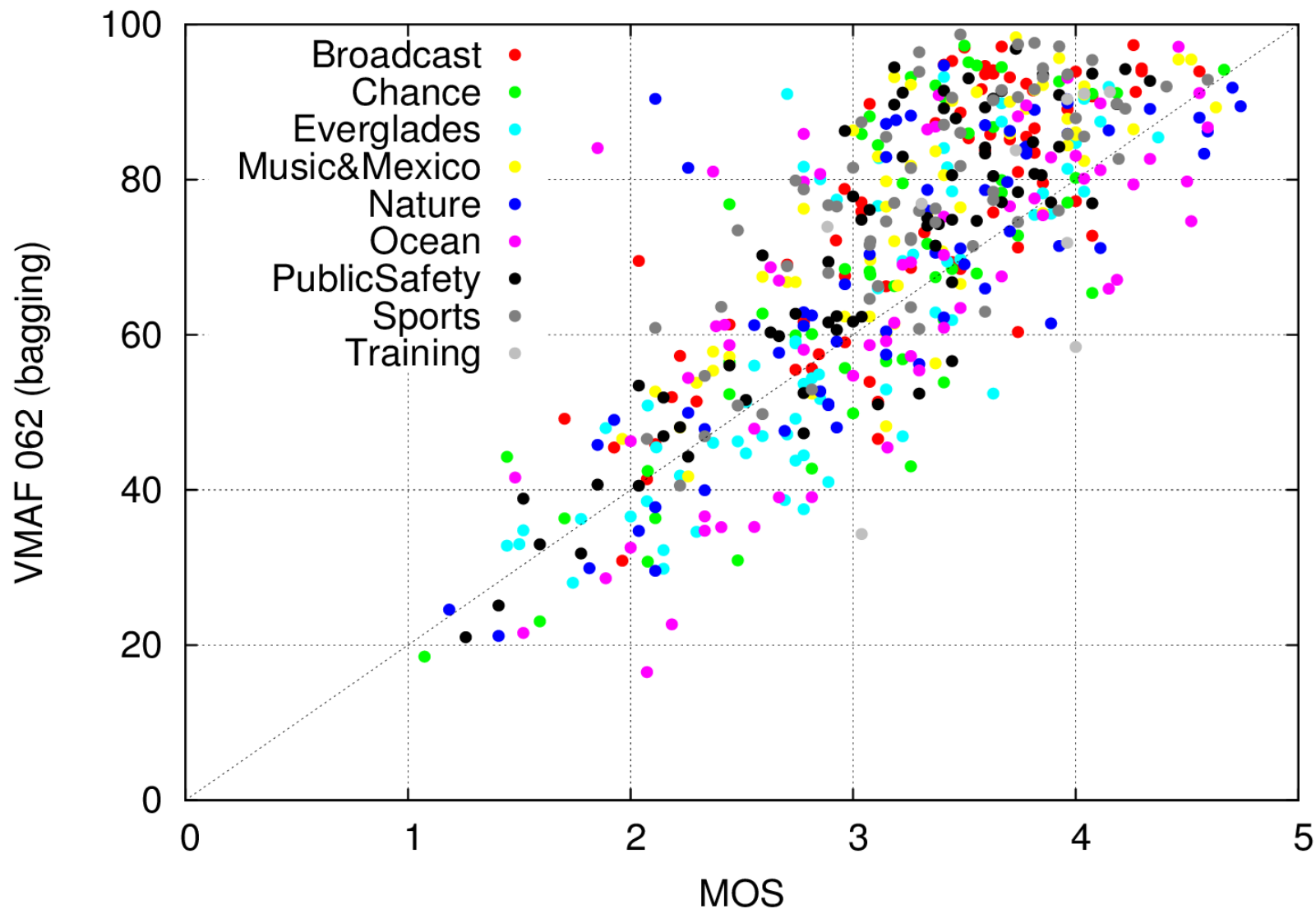


# Results: VMAF 0.6.2 vs MOS

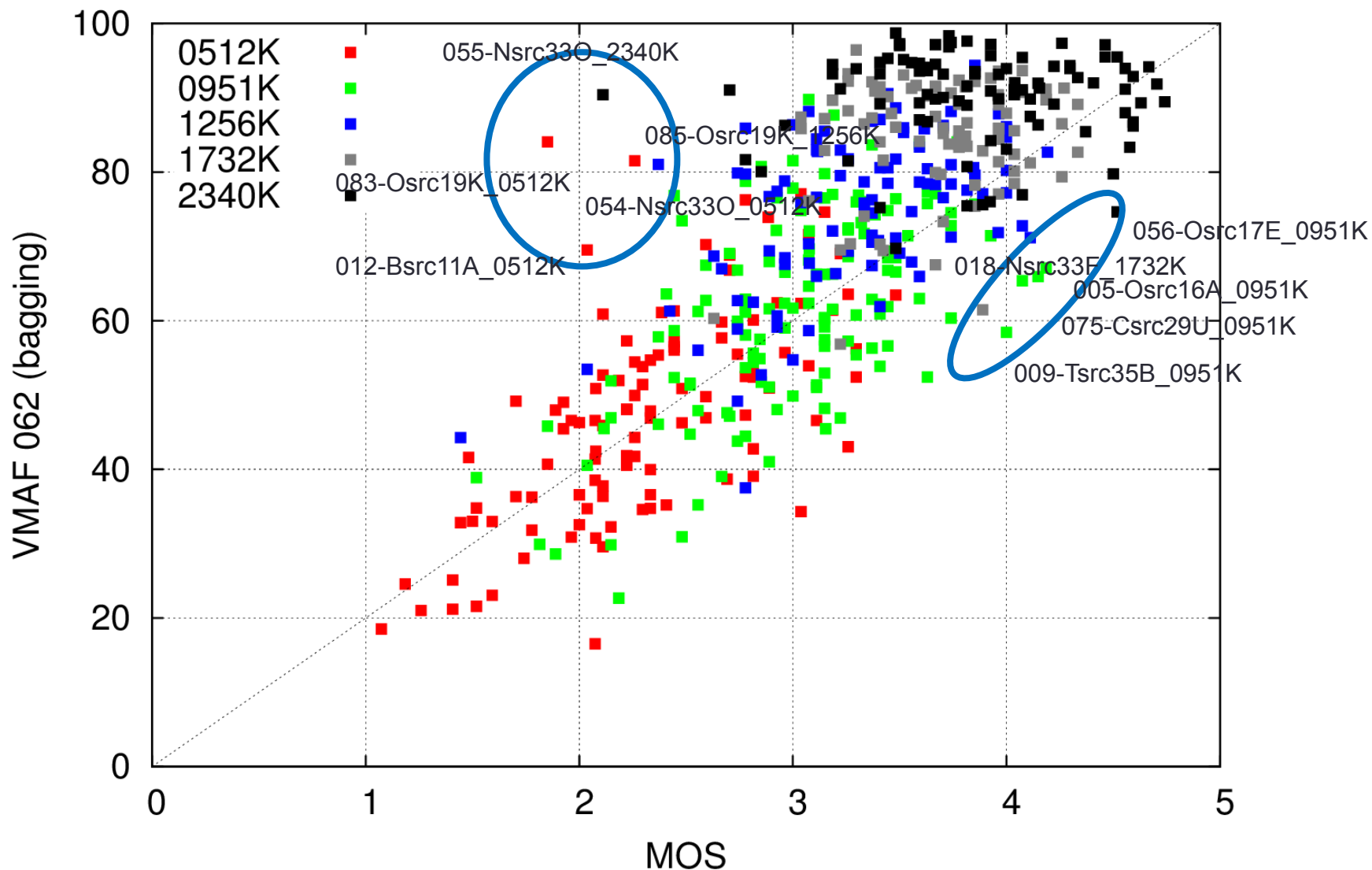


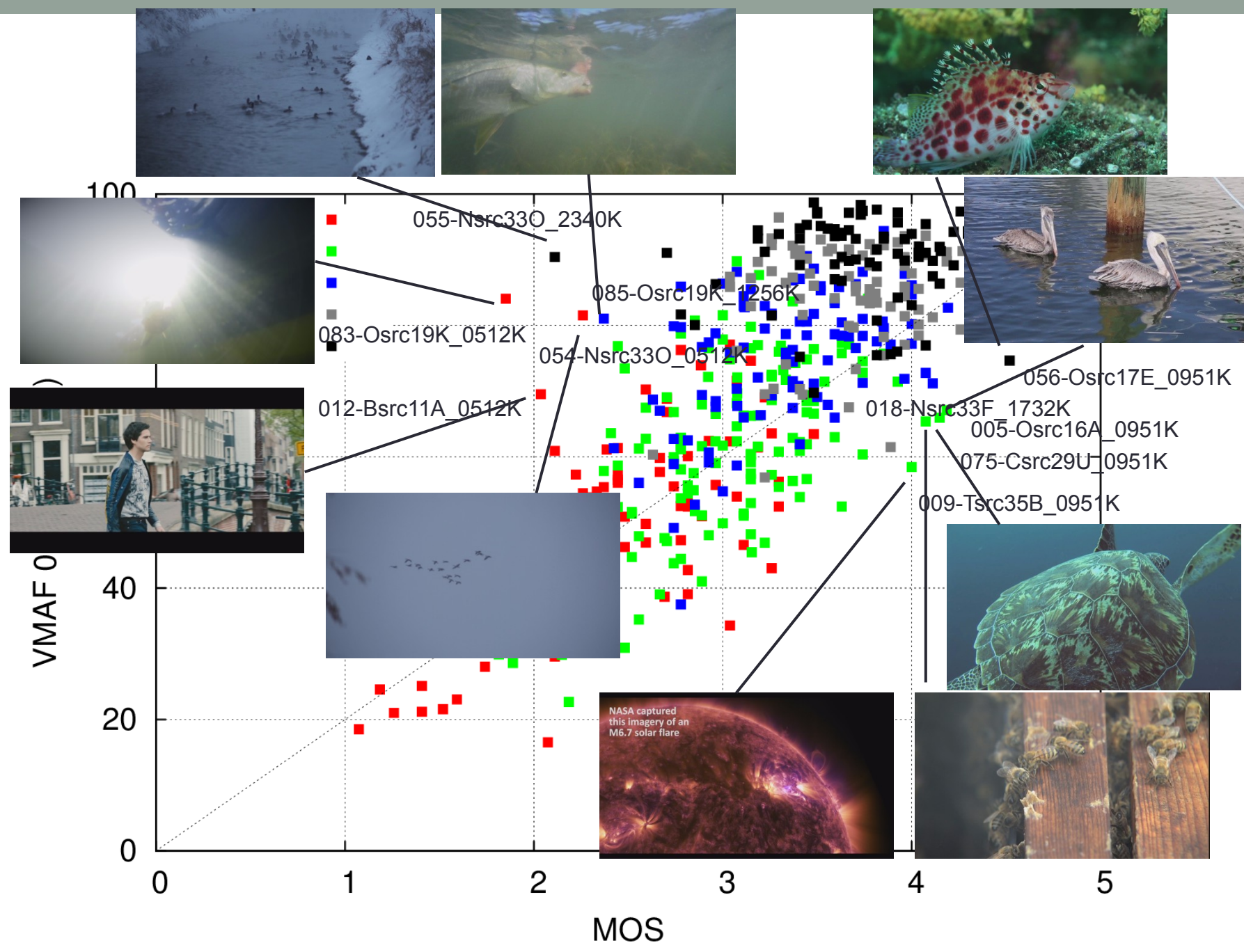


# Results: VMAF 0.6.2 vs MOS



# Results: VMAF 0.6.2 vs MOS





## Visual Comparison (actual MOS < predicted)



## Visual Comparison (actual MOS < predicted)



## Visual Comparison (actual MOS < predicted)



## Visual Comparison (actual MOS < predicted)



## Visual Comparison (actual MOS < predicted)





# Visual Comparison (actual MOS < predicted)



## Visual Comparison (actual MOS > predicted)



## Visual Comparison (actual MOS > predicted)



## Visual Comparison (actual MOS > predicted)



## Visual Comparison (actual MOS > predicted)



## Visual Comparison (actual MOS > predicted)

NASA captured  
this imagery of an  
M6.7 solar flare



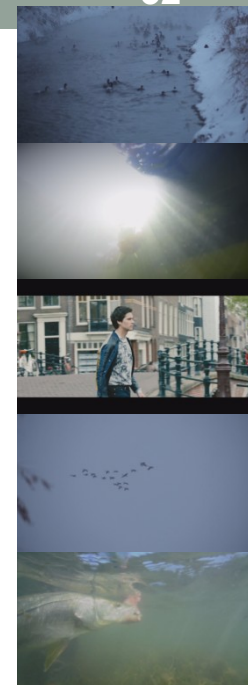
## Visual Comparison (actual MOS > predicted)

NASA captured  
this imagery of an  
M6.7 solar flare

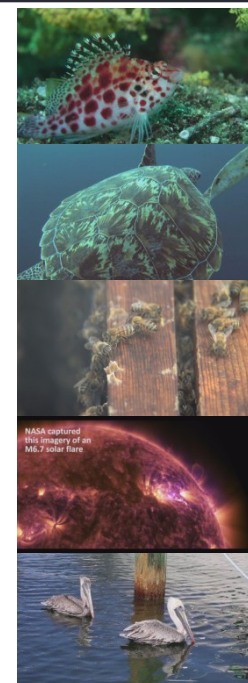


# Objective Measures

name	PSNR	SSIM	MS-SSIM	VIF	VMAF	VMAF	VMAF	MOS	MOS	bitrate	session
					v.0.6.2	b.aggreg	CI	value	CI		
055-Nsrc33O_2340K	46,02	0,994	0,995	0,731	91,878	90,411	5,045	2,111	0,556	2340000	N
083-Osrc19K_0512K	43,84	0,994	0,995	0,661	83,154	84,063	6,687	1,852	0,606	512000	O
012-Bsrc11A_0512K	<b>30,13</b>	0,962	0,969	<b>0,466</b>	69,916	69,522	6,636	2,037	0,435	512000	B
054-Nsrc33O_0512K	41,81	0,993	0,994	0,568	79,990	81,507	6,076	2,259	0,602	512000	N
085-Osrc19K_1256K	40,28	0,990	0,992	0,602	80,817	81,035	4,930	2,370	0,586	1256000	O
056-Osrc17E_0951K	31,80	0,979	0,980	0,570	66,432	67,084	3,948	4,185	0,504	951000	O
005-Osrc16A_0951K	32,58	0,979	0,981	<b>0,354</b>	64,728	65,924	3,223	4,148	0,490	951000	O
075-Csrc29U_0951K	34,78	0,973	0,976	0,514	65,197	65,366	3,692	4,074	0,540	951000	C
009-Tsrc35B_0951K	31,58	0,974	0,976	0,494	56,333	58,435	3,922	4,000	0,587	951000	T
018-Nsrc33F_1732K	29,89	0,937	0,944	<b>0,386</b>	63,703	61,466	6,095	3,889	0,660	1732000	N



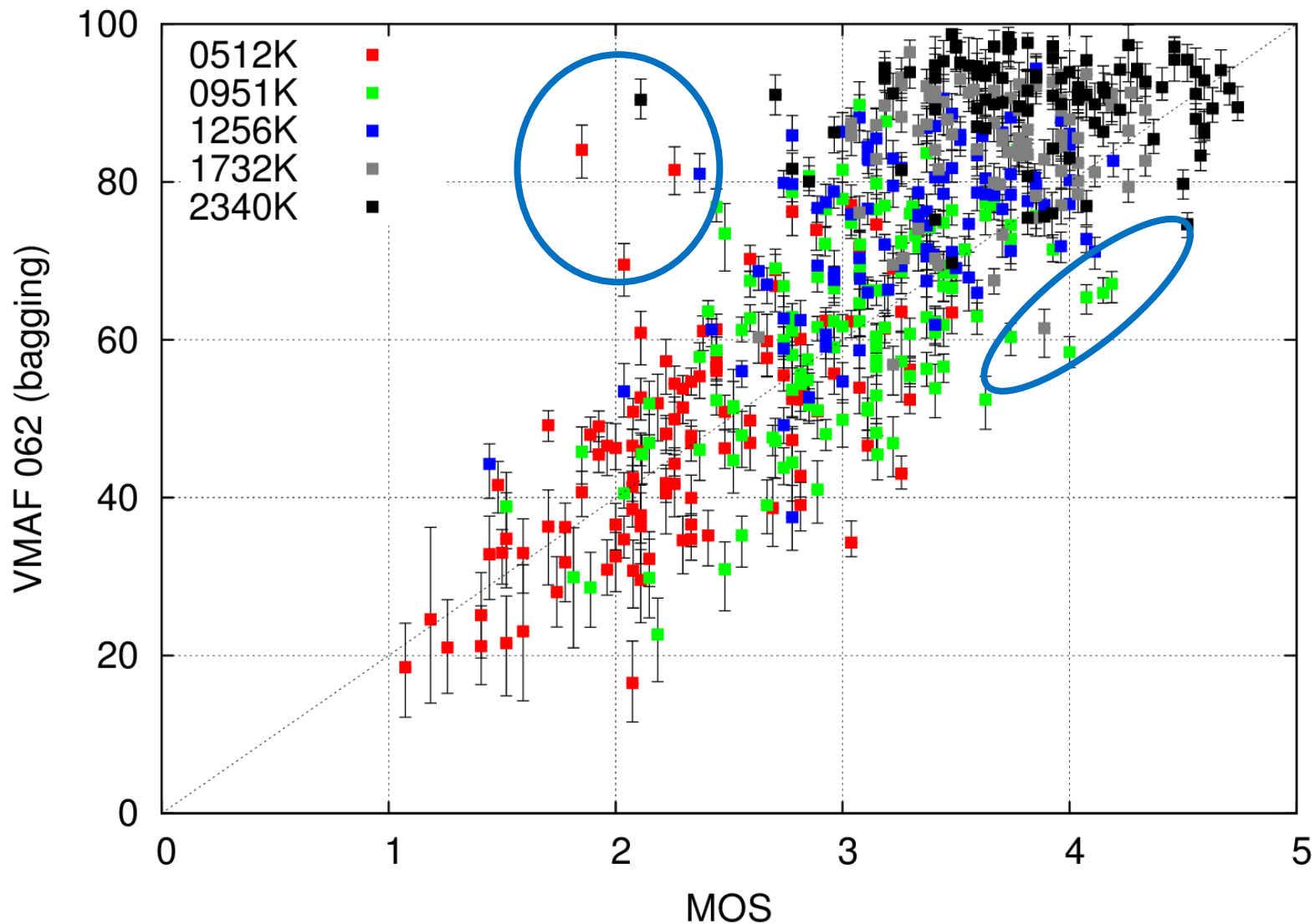
Low MOS



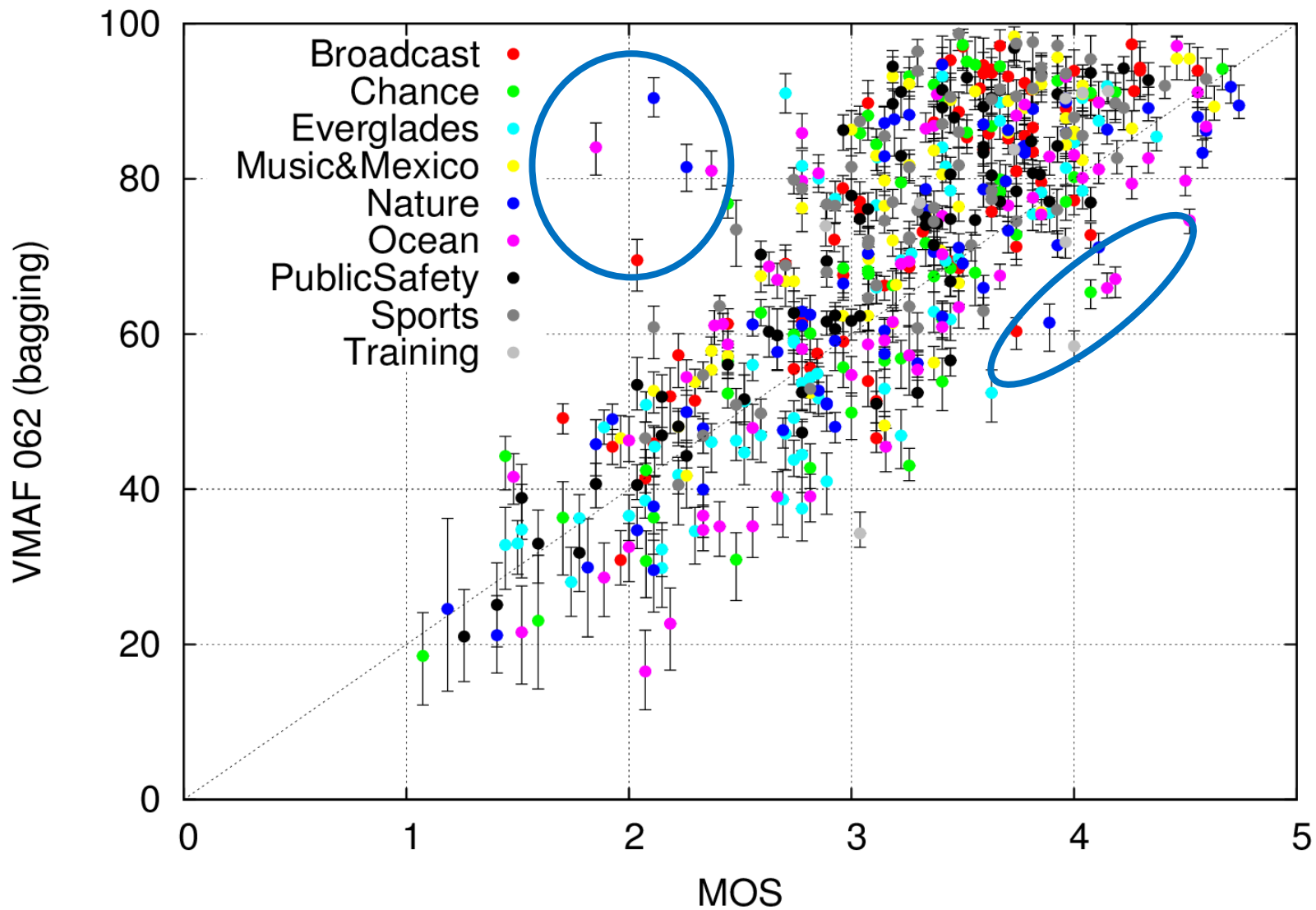
High MOS



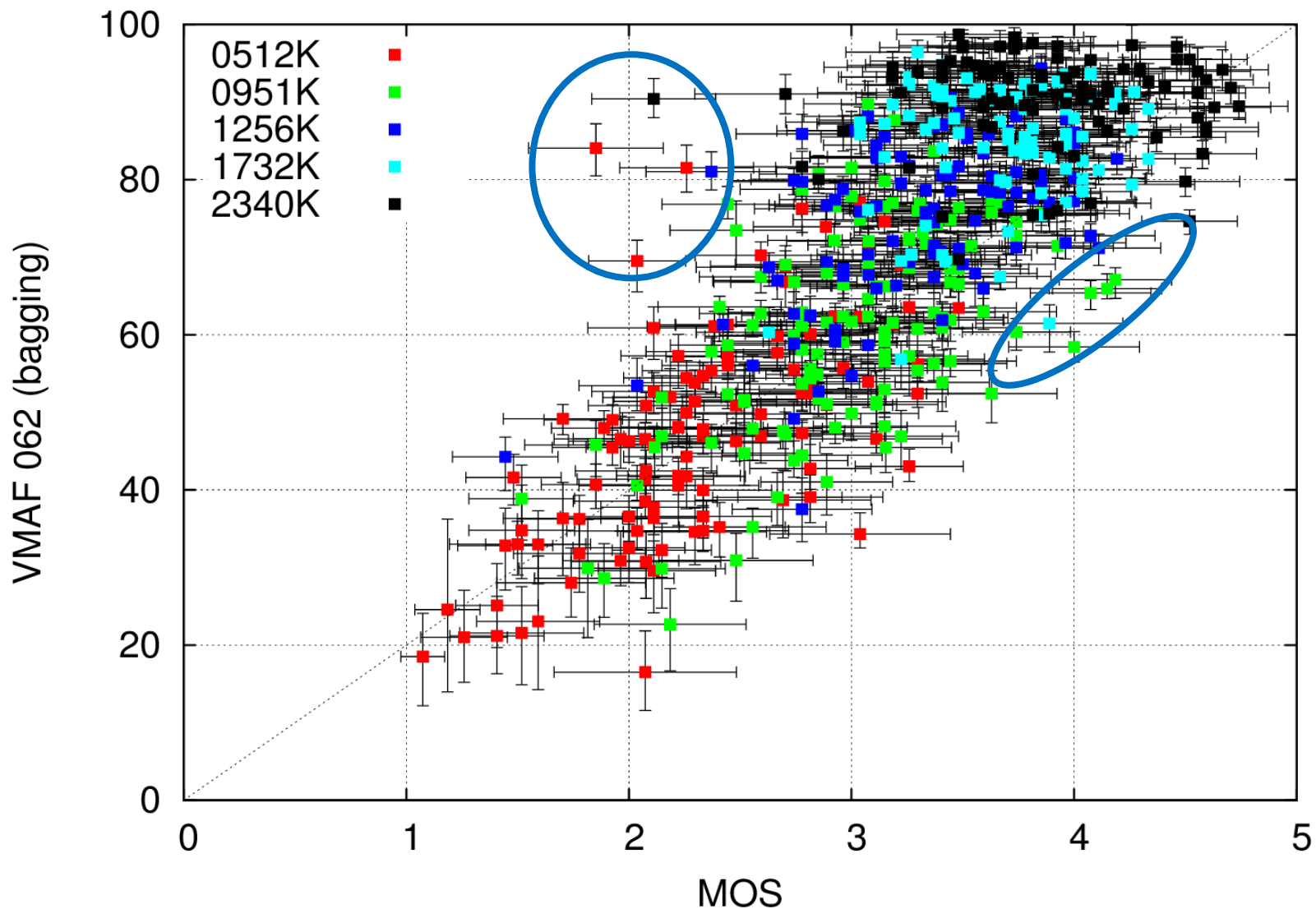
# Results: VMAF 0.6.2 (CI) vs MOS



# Content Dependency

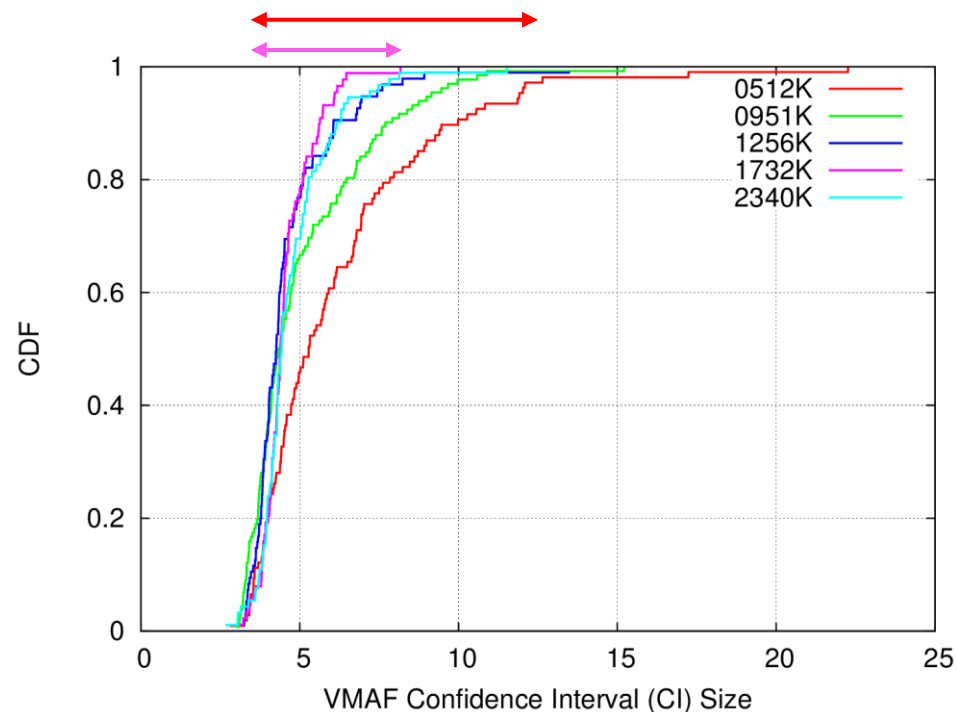
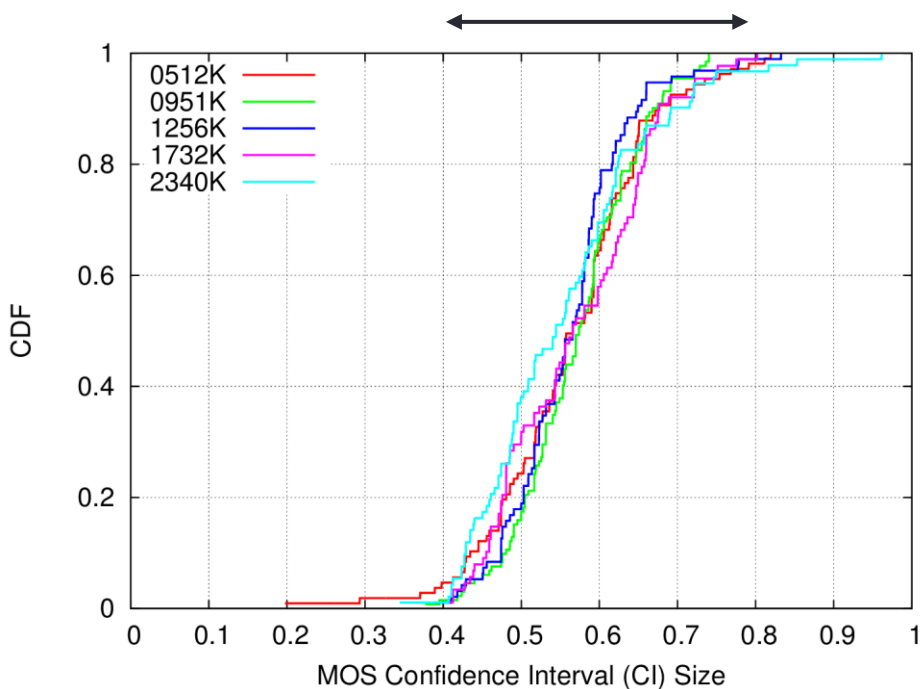


# VMAF & MOS CIs (95%)



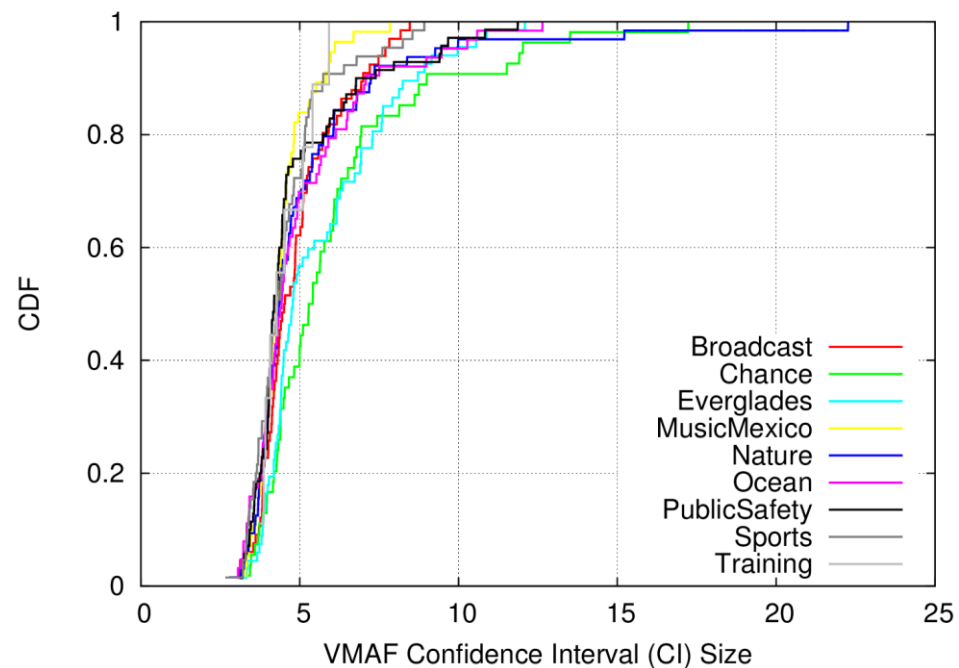
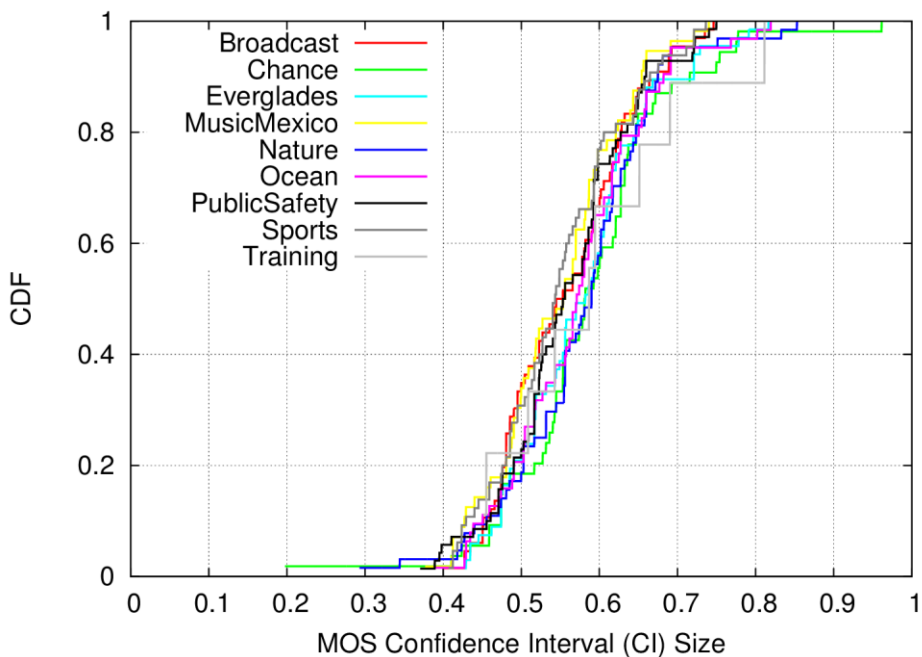
# Confidence Intervals: MOS & VMAF

- As a function of the encoding bitrate
- MOS: about [0.4-0.8]
- VMAF:  $[3-12]/100 = [0.15-0.6]/5$ , higher for lower bitrates



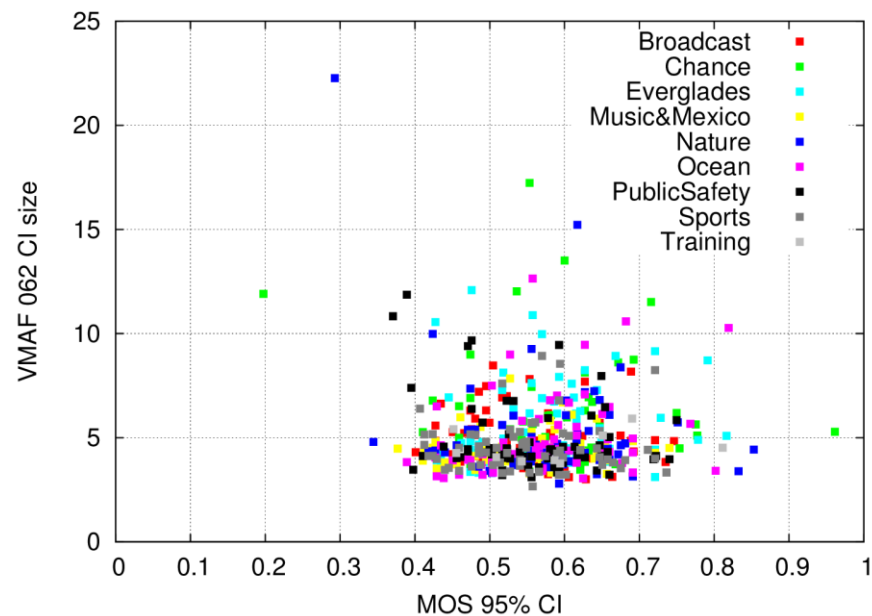
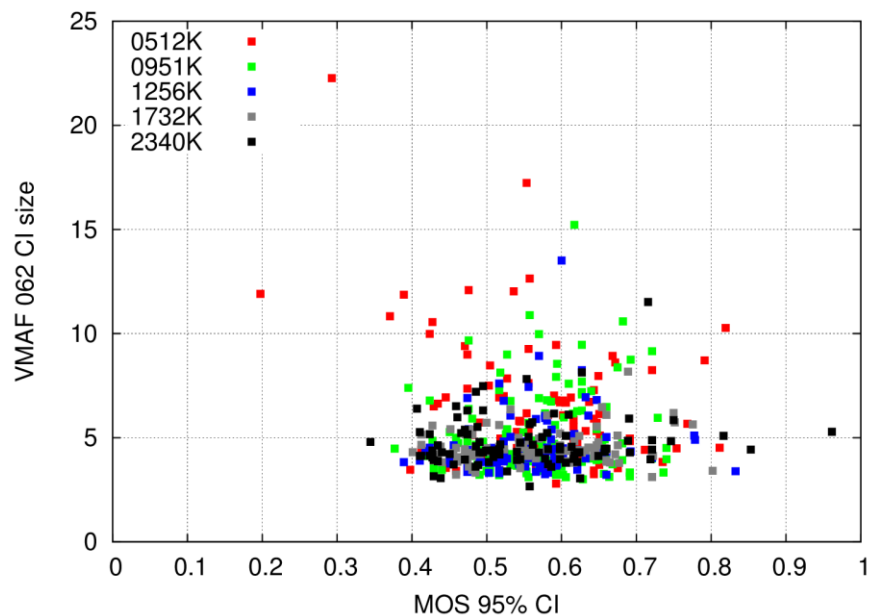
# Confidence Intervals: MOS & VMAF

- As a function of the session type

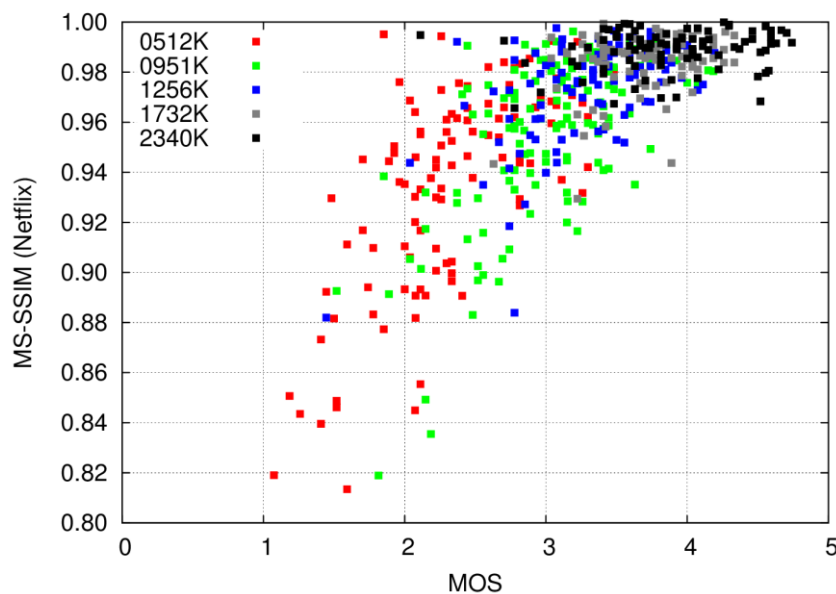
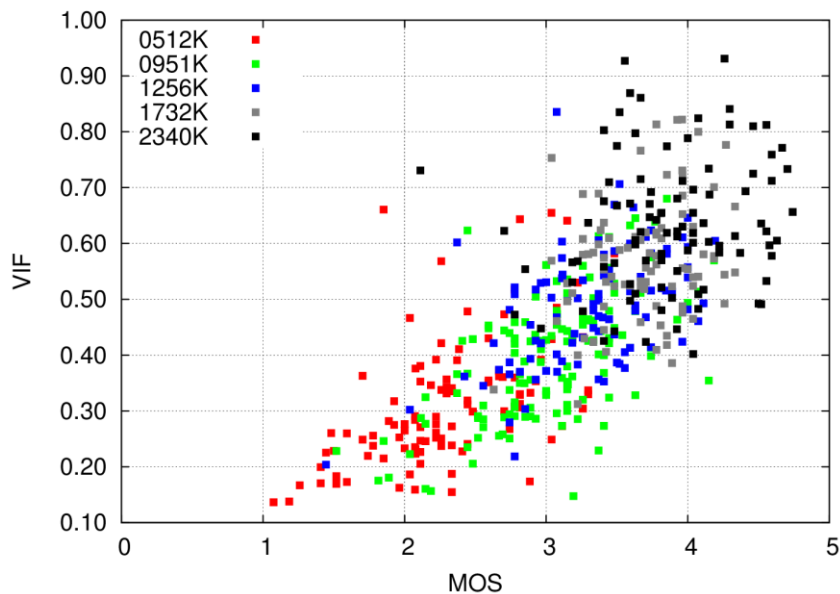
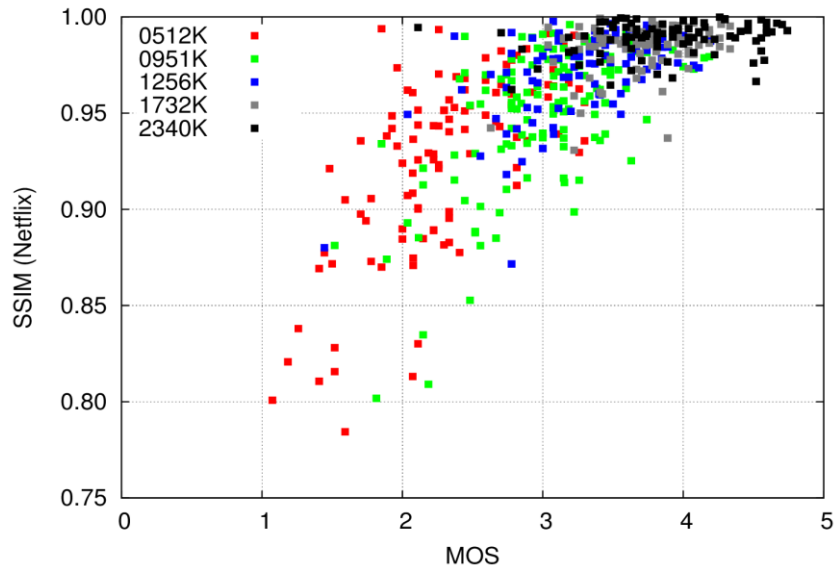
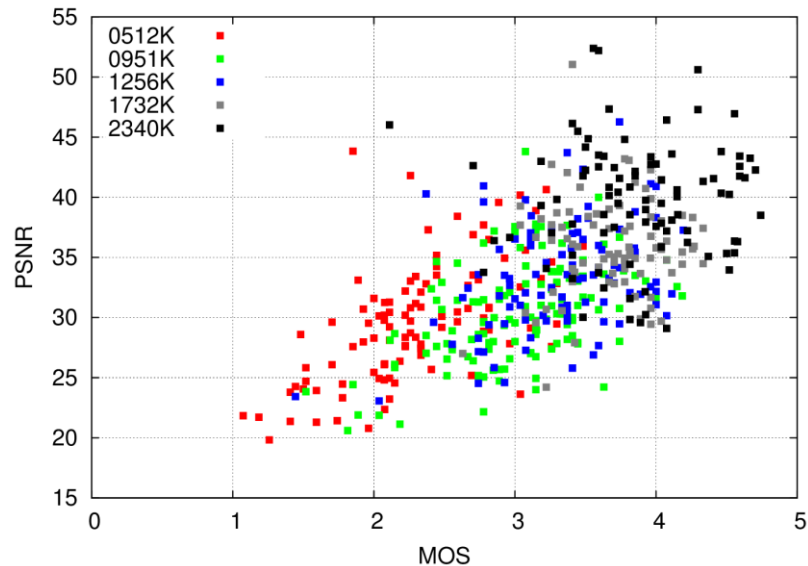


# Confidence Intervals: MOS & VMAF

- Scatter plot: no particular behavior

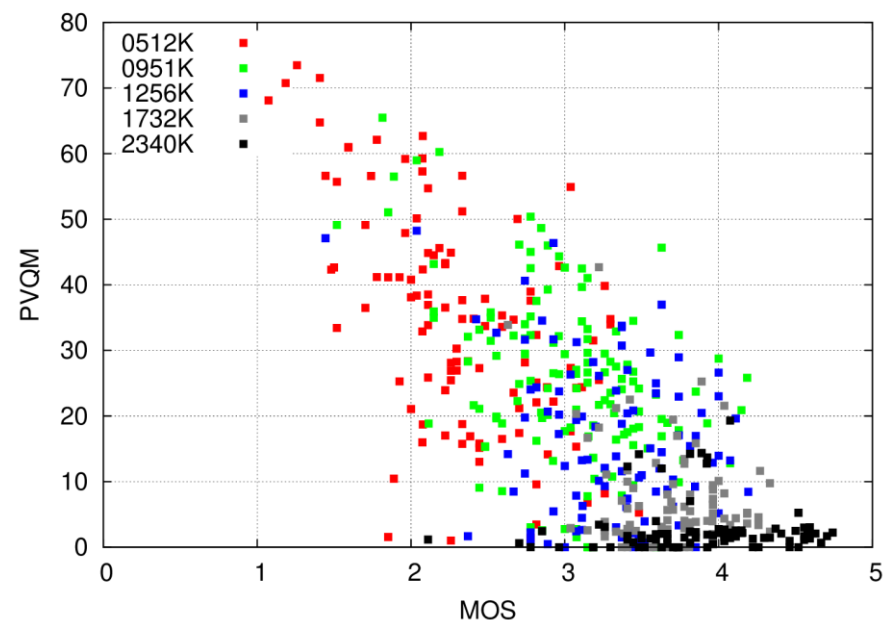
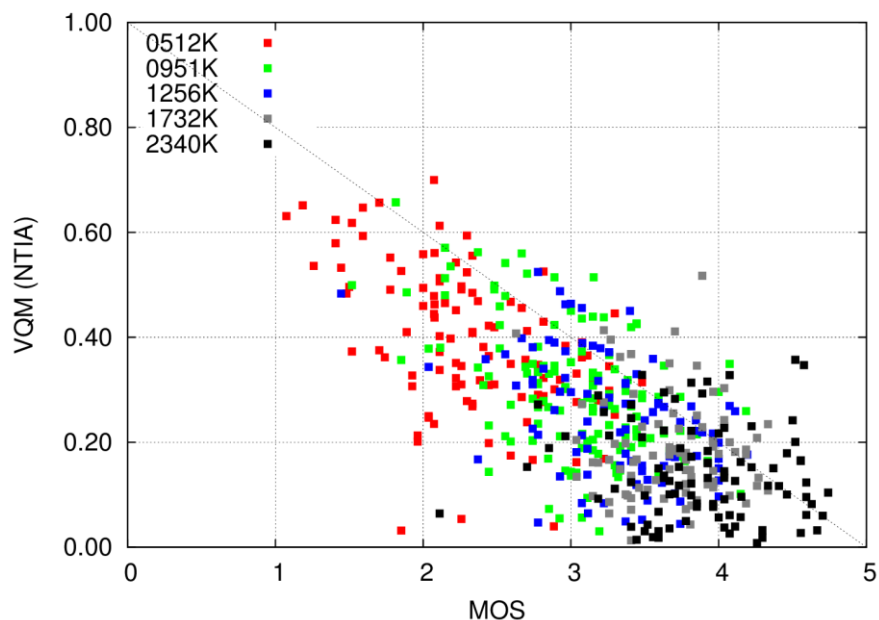


# Other Measures



# Other Measures

- VQM, PVQM (not a simple arithmetic mean of frame-based measures)

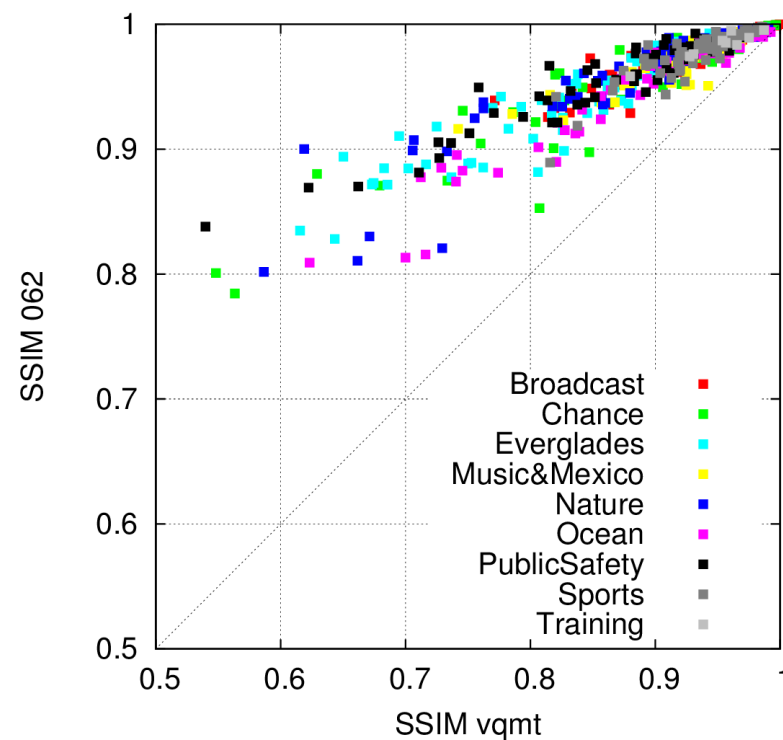
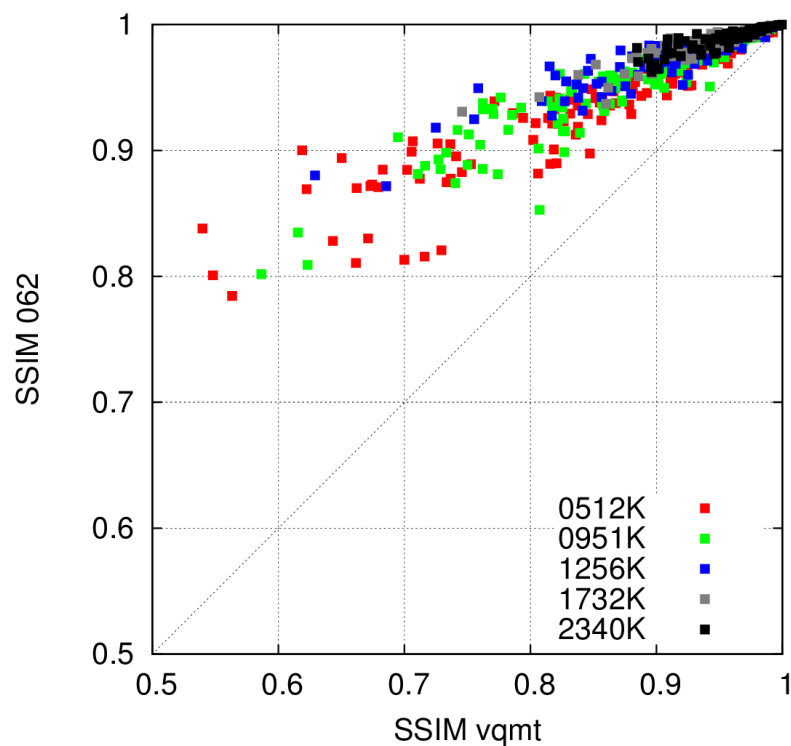


- VQM: ITU-T J.144, "Objective perceptual video quality measurement techniques for digital cable television in the presence of a full reference", 2001.
- PVQM: Hekstra AP, Beerends JG, Ledermann D, De Caluwe FE, Kohler S, Koenen RH, Rihs S, Ehram M, Schlauss D., "PVQM—A perceptual video quality measure", Signal processing: Image communication. 2002 Nov 1;17(10):781-98.



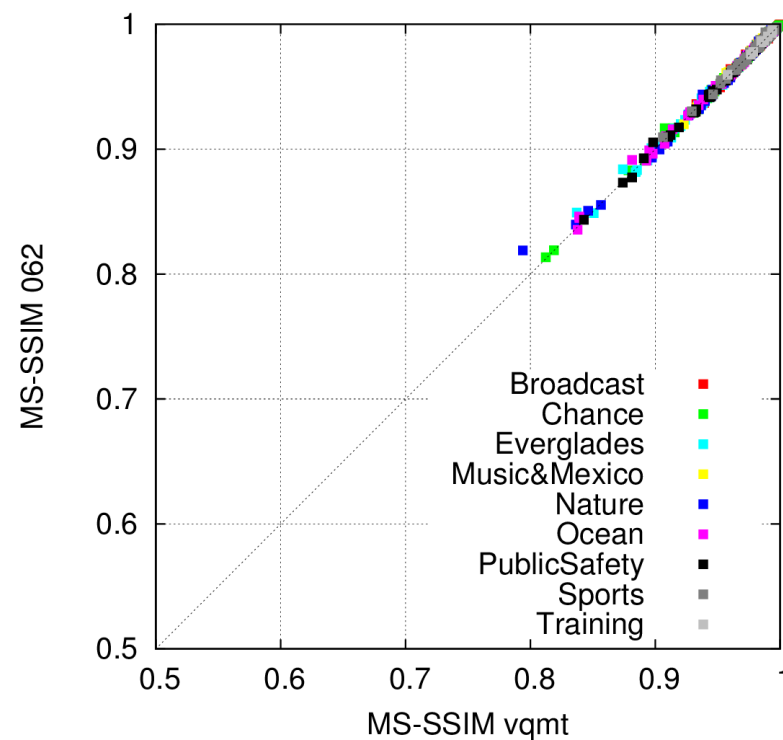
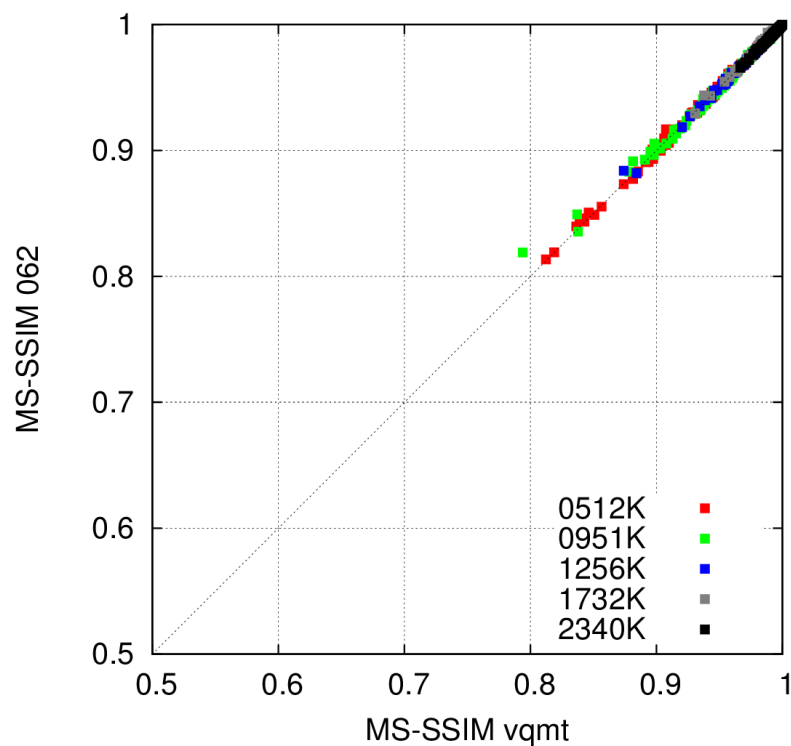
# SSIM: Netflix vs vqmt

- Same C1 and C2 constants in the formula
- SSIM vs downsampled SSIM?



# MS-SSIM: Netflix vs vqmt

- Only small differences in case of MS-SSIM



# Results

- Data available here:
  - CSV file: <http://media.polito.it/downloads/jeg/its4s/>
  - On Google Datastudio: [https://bit.ly/its4s\\_2018](https://bit.ly/its4s_2018)  
[https://datastudio.google.com/open/1MUqys1gsnEdKJHFK5q\\_112oEYqB9e4xZ](https://datastudio.google.com/open/1MUqys1gsnEdKJHFK5q_112oEYqB9e4xZ)



# Conclusions

- Large content variety helps in characterizing objective measures and find areas of improvement
- Confidence intervals by VMAF might be useful, but bootstrapping aggregation value should be used
- Size of confidence intervals in subjective experiments is mostly between [0.4-0.8] MOS
- Size of VMAF confidence interval is mostly between  $[3-12]/100 = [0.15-0.6]/5$ , with higher values for lower bitrates

# Future Plans

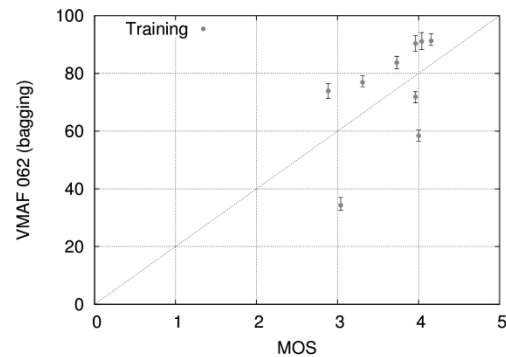
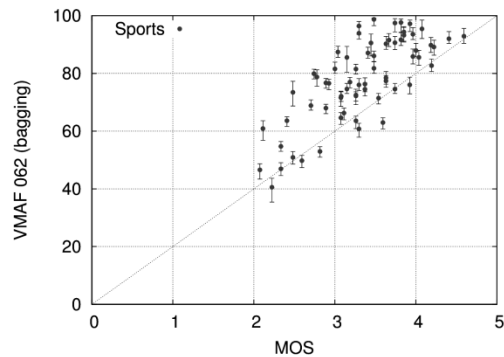
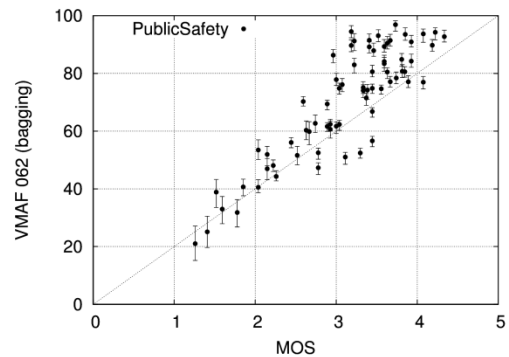
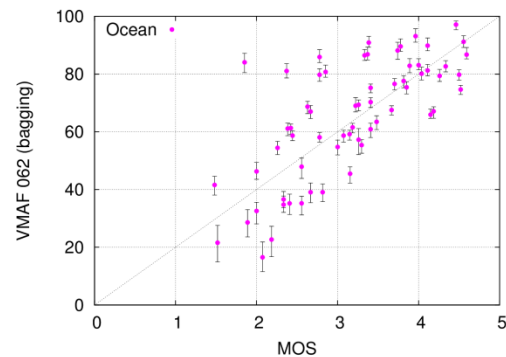
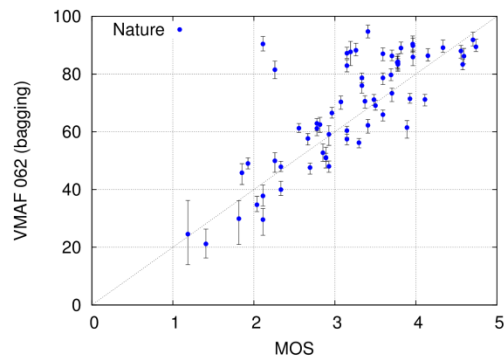
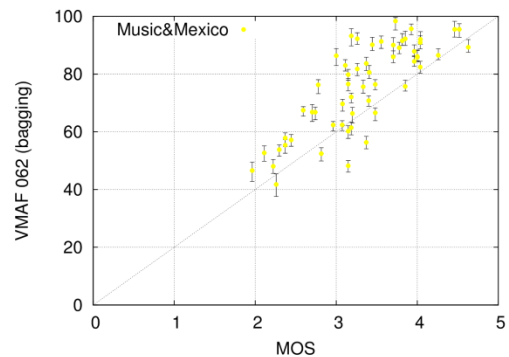
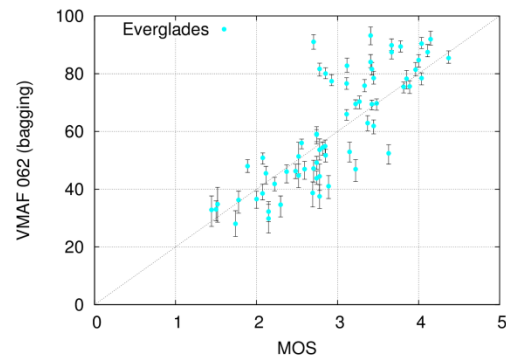
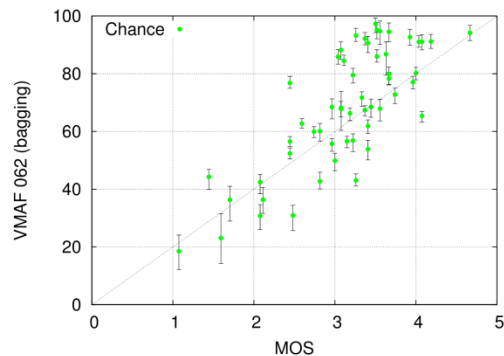
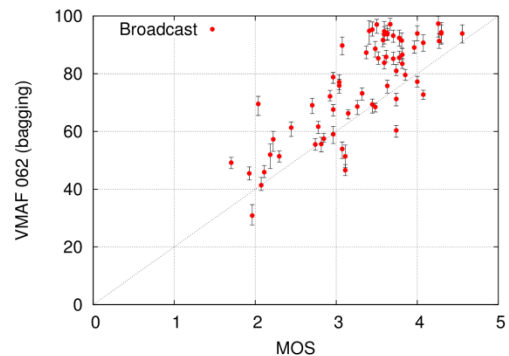
- Better characterize observations that can be done with objective measures, trying to understand how much they can predict potentially difficult situations in subjective evaluation
- Compute the new objective measures on the Large-Scale JEG-Hybrid database of ~60,000 sequences
  - VMAF 0.6.2 with confidence intervals
  - SSIM and MS-SSIM by Netflix (and compare them with the vqmt)

# References

- M. Pinson, “ITS4S: A Video Quality Dataset with Four-Second Unrepeated Scenes”, TM-18-532, 2018
- ITU-T J.144, "Objective perceptual video quality measurement techniques for digital cable television in the presence of a full reference", 2001.
- Hekstra AP, Beerends JG, Ledermann D, De Caluwe FE, Kohler S, Koenen RH, Rihs S, Ehram M, Schlauss D., “PVQM—A perceptual video quality measure”, Signal processing: Image communication. 2002 Nov 1;17(10):781-98.
- M. Barkowsky, E. Masala, G. Van Wallendael, K. Brunnstrom, N. Staelens, P. Le Callet, Objective Video Quality Assessment – Towards large scale video database enhanced model development, IEICE Transactions on Communications, vol. E98-B, n. 1, pp. 2-11, Jan 2015
- [ftp://iotnas001.th-deg.de/VQEG/JEG/HYBRID/hevc\\_database/](ftp://iotnas001.th-deg.de/VQEG/JEG/HYBRID/hevc_database/)
- <http://media.polito.it/downloads/jeg/>

- Backup slides

# Content Dependency





# Content Dependency

