On the suitability of VMAF for quality assessment of medical videos: Medical ultrasound & wireless capsule endoscopy

Arslan Usman and Maria Martini

Presenter:
Prof Maria Martini
Wireless Multimedia and Networking Research Group
Kingston University, London, UK
What makes it more challenging?

- In medical images the diagnostic quality of the image is more important than the perceptual quality.
- However, typical objective image quality metrics do not measure the diagnostic quality.
Option 1: develop tailored metrics

Examples:

• A Raj, NA Shah, AK Tiwari, MG Martini, Multivariate Regression-Based Convolutional Neural Network Model for Fundus Image Quality Assessment, *IEEE Access* 8, 2020, 57810-57

• M Razaak and MG Martini, "CUQI: Cardiac Ultrasound Quality Index," *SPIE Journal of Medical Imaging*, 2016.
Option 2: study which of the existing metrics performs best

PSNR, SSIM, UQI, VQM, NQM, VIF, NIQE, BRISQUE were tested earlier for Ultrasound video, Wireless Capsule Endoscopy, ENT Endoscopic Video, Radiological images

- M. Razaak, M.G. Martini and K. Savino, "A Study on Quality Assessment for Medical Ultrasound Video Compressed via HEVC," *IEEE Journal of Biomedical and Health Informatics (J-BHI)*, vol. 18, no. 5, pp. 1552-1559, Sep 2014.

- Usman, M. A., Usman, M. R., & Shin, S. Y. (2017). Quality assessment for wireless capsule endoscopy videos compressed via HEVC: From diagnostic quality to visual perception. *Computers in biology and medicine, 91*, 112-134.

- Chaabouni, A., Gaudeau, Y., Lambert, J., Moureaux, J.M. and Gallet, P., 2014, October. Subjective and objective quality assessment for H264 compressed medical video sequences. In 2014 4th International Conference on Image Processing Theory, Tools and Applications (IPTA) (pp. 1-5). IEEE.

- Kowalik-Urbaniak, Ilona, Dominique Brunet, Jiheng Wang, David Koff, Nadine Smolarski-Koff, Edward R. Vrscay, Bill Wallace, and Zhou Wang. "The quest for 'diagnostically lossless' medical image compression: a comparative study of objective quality metrics for compressed medical images." In *Medical Imaging 2014: Image Perception, Observer Performance, and Technology Assessment*, vol. 9037, p. 903717. International Society for Optics and Photonics, 2014.
Option 2: study which of the existing metrics performs best, ctd.

VMAF was developed for and trained on non-medical video, but has shown excellent performance for different types of video (e.g. gaming)

Challenge: what if we apply it to medical video?
Ultrasound video dataset

- Nine ultrasound video sequences provided by cardiologist in Perugia Hospital (Heart, Liver, Kidney, Lung), 640 x 416, 100 frames, 25fps
- Compressed with HEVC at eight different compression ratios

9 x 8 = 72 video sequences being evaluated with DSCQS (144 sequences evaluated by each specialist not including intra-subject controls)

[Razaak, Martini, Savino, IEEE JBHI 2014]

Extensive subjective tests run with medical doctors in Hospital of Perugia
Ten source video sequences provided by Intromedic Co. Ltd., South Korea, for research purposes.

3 fps (Native)
320 x 320 (Native)
10 s

HEVC compression with QP 27, 29, 31, 33, 35, 37, 39 & 41
# Medical Image quality evaluation

Other Objective Quality Metrics considered

| Quality metric                                      | Abbreviation |
|----------------------------------------------------|--------------|
| Peak Signal to Noise Ratio                         | PSNR         |
| Structural Similarity Index Metric [13]            | SSIM         |
| Multi Scale SSIM [14]                              | MS-SSIM      |
| Visual Signal to Noise Ratio [15]                  | VSNR         |
| Information Fidelity Criterion [16]                | IPC          |
| Visual Information Fidelity [17]                   | VIF          |
| Pixel-based VIF [17]                               | VIFP         |
| Universal Quality Index [18]                       | UQI          |
| Noise Quality Measure [19]                         | NQM          |
| Weighted Signal to Noise Ratio [19]                | WSNR         |
| Video Quality Metric [20]                          | VQA<sub>STD</sub> |
| Video Multimethod Assessment Fusion [10]           | VMAF         |
Results - 1

Accuracy comparison

Results for fitting the VMAF measurements to the subjective DMOS.

| Dataset       | Category   | Exponential | Linear | Logistic |
|---------------|------------|-------------|--------|----------|
|               |            | R² | Adj. R² | RMSE | R² | Adj. R² | RMSE | R² | Adj. R² | RMSE |
| Ultrasound Videos | Expert | 0.8032 | 0.8004 | 11.1736 | 0.8620 | 0.8601 | 9.3587 | 0.8544 | 0.8502 | 9.680 |
|               | Non-Expert | 0.8334 | 0.8310 | 8.7630 | 0.8862 | 0.8846 | 7.242 | 0.8791 | 0.8756 | 7.5199 |
| WCE Videos    | Expert | 0.9214 | 0.9204 | 3.8859 | 0.9267 | 0.9258 | 4.9820 | 0.9268 | 0.9239 | 3.7994 |
|               | Non-Expert | 0.9501 | 0.9494 | 3.0967 | 0.9433 | 0.9426 | 5.8556 | 0.9501 | 0.9481 | 3.1370 |

Comparison of VMAF with other FR-VQMs.

| Dataset | Scores | CC | PSNR | SSIM | MS-SSIM | VSNR | WSNR | NQM | UQI | VIF | VIP | IFC | VQM NISTA | VMAF |
|---------|--------|----|------|------|---------|------|------|-----|-----|-----|-----|-----|----------|------|
| Ultra-sound | Experts | PLCC | 0.9109 | 0.9254 | 0.8579 | 0.8925 | 0.9123 | 0.8961 | 0.9292 | 0.9258 | 0.8887 | 0.8644 | 0.8080 | 0.9056 |
|          | Non-experts | PLCC | 0.9331 | 0.9375 | 0.8907 | 0.9139 | 0.9251 | 0.9090 | 0.9251 | 0.9382 | 0.8997 | 0.8926 | 0.8368 | 0.8941 |
| WCE      | Experts | PLCC | 0.9280 | 0.9383 | 0.8899 | 0.9277 | 0.9354 | 0.9464 | 0.9495 | 0.9668 | 0.9047 | 0.8906 | 0.8606 | 0.9186 |
|          | Non-experts | PLCC | 0.8039 | 0.6840 | 0.8566 | 0.6055 | 0.8010 | 0.7158 | 0.8701 | 0.9016 | 0.8055 | 0.8844 | 0.7764 | 0.9627 |
|          | PLCC | 0.8611 | 0.8653 | 0.9127 | 0.6571 | 0.8709 | 0.8257 | 0.8930 | 0.9424 | 0.9263 | 0.9482 | 0.8426 | 0.9763 |
|          | SROCC | 0.8257 | 0.7232 | 0.8696 | 0.6204 | 0.7963 | 0.7371 | 0.8909 | 0.9238 | 0.9227 | 0.9620 | 0.7578 | 0.9712 |
|          | PLCC | 0.8642 | 0.8129 | 0.9247 | 0.6474 | 0.8774 | 0.8311 | 0.9061 | 0.9533 | 0.9408 | 0.9525 | 0.8402 | 0.9796 |
Results - 2

Scatter plots
Observations and proposed future steps

- Limited number of scores from experts in available datasets
- The level of expertise of the “experts” influences the quality scores (not all expert subjects are equal!)
- More datasets with videos assessed by a wide range of experts are required
References

A Raj, NA Shah, AK Tiwari, MG Martini, Multivariate Regression-Based Convolutional Neural Network Model for Fundus Image Quality Assessment, *IEEE Access* 8, 2020, 57810-57

Raj, A K Tiwari, MG Martini, Fundus image quality assessment: survey, challenges, and future scope, *IET Image Processing* 13 (8), 2019. 1211-1224

Usman, M. A., & Martini, M. G. (2019). On the suitability of VMAF for quality assessment of medical videos: Medical ultrasound & wireless capsule endoscopy. *Computers in biology and medicine*, 113, 103383.

Lévêque, L., Liu, H., Baraković, S., Husić, J. B., Martini, M., Outtas, M., ... & Pinheiro, A. (2018, May). On the subjective assessment of the perceived quality of medical images and videos. In *2018 Tenth International Conference on Quality of Multimedia Experience (QoMEX)* (pp. 1-6). IEEE.

Kara, P. A., Kovacs, P. T., Vagharshakyan, S., Martini, M. G., Imre, S., Barsi, A., ... & Balogh, T. (2017). Perceptual quality of reconstructed medical images on projection-based light field displays. In *eHealth 360°* (pp. 476-483). Springer, Cham.

Usman, M. A., Usman, M. R., & Shin, S. Y. (2017). Quality assessment for wireless capsule endoscopy videos compressed via HEVC: From diagnostic quality to visual perception. *Computers in biology and medicine*, 91, 112-134.

Nasr, K.M. and Martini, M.G., 2017. A visual quality evaluation method for telemedicine applications. *Signal Processing: Image Communication*, 57, pp.211-218.

M. Razaak and M.G. Martini, "CUQI: Cardiac Ultrasound Quality Index," *SPIE Journal of Medical Imaging*, 2016.

M. Razaak, M.G. Martini and K. Savino, "A Study on Quality Assessment for Medical Ultrasound Video Compressed via HEVC," *IEEE Journal of Biomedical and Health Informatics (J-BHI)*, vol. 18, no. 5, pp. 1552-1559, Sep 2014.
Thank you!

On the suitability of VMAF for quality assessment of medical videos: Medical ultrasound & wireless capsule endoscopy

Arslan Usman and Maria Martini

Presenter: Prof Maria Martini
Wireless Multimedia and Networking Research Group
Kingston University, London, UK