# NoVisIQ: Convolutional Neural Network for No-Reference Objective Image Quality Assessment

Kylie Cancilla, Alexander Moore, Amar Saini

Video object detection often struggles with challenges like blur, low lighting, and poor network quality. To address this, we developed a no-reference deep learning model trained on synthetically distorted data and labeled with objective full-reference metrics. This model quantifies frame quality without a reference, supporting more robust video object detection in difficult conditions.

## INTRODUCTION

Video is everywhere today, from smartphones and security cameras to advanced technologies like autonomous laboratories and self-driving cars. In these applications, object detection in video is especially powerful because it adds temporal context, making it possible to recognize complex behaviors and patterns over time. This enables real-time responses, such as alerting drivers to hazards or allowing robotic arms to adjust their movements on the fly. However, real-world video comes with real-world challenges: noise, motion blur, low light, and network interruptions can all degrade video quality and, in turn, reduce the accuracy of object detection systems.

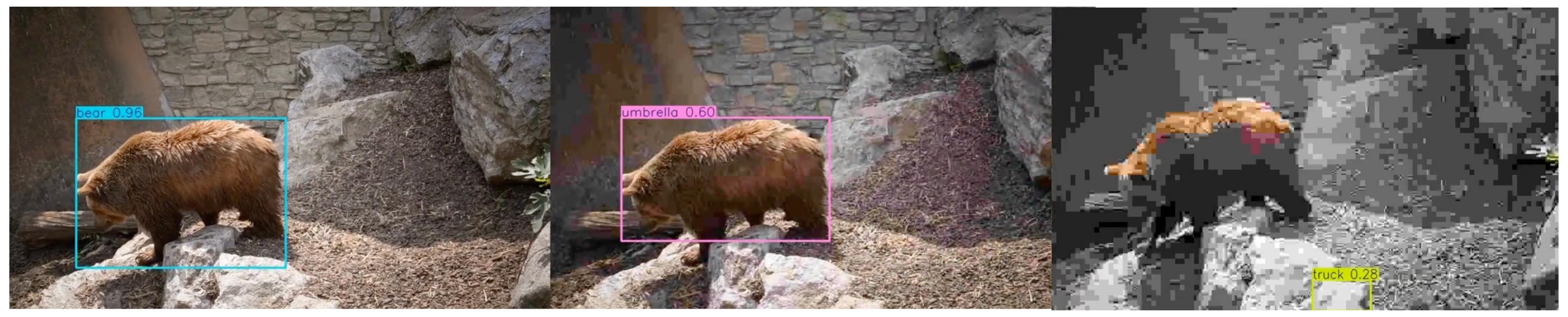

***Figure 1:** After JPEG compression, the object detection model fails to detect a bear, even though the bear remains clearly visible to the human eye.*

Assessing video quality is crucial for maintaining reliable performance, but most existing image quality metrics have significant limitations. Full-reference metrics such as LPIPS [1], PSNR [2], and SSIM [3] require access to original, unaltered images, which are rarely available in real-world scenarios. No-reference metrics like BRISQUE and BLINDS [4] rely on costly, subjective human labels, making them impractical for large-scale or automated applications. To address these challenges, we have developed a no-reference, objective image quality assessment method that can quantify the quality of video frames efficiently. This approach provides actionable insights to improve object detection performance in challenging video environments.

## METHODS

To overcome the limitations of existing metrics, our model is trained on images that have been augmented with various distortions, including blur, dimming, noise, and compression. Full-reference metrics are used to generate synthetic labels during training, enabling the model to learn to predict image quality without needing a reference at inference time. We built two convolutional neural networks (CNNs) from scratch and applied transfer learning to pre-trained architectures. All models were trained and evaluated to determine which performed best for the task.

$$\text{Total Loss} = \frac{1}{3}\left(\text{MAE}_{\text{LPIPS}} + \text{MAE}_{\text{PSNR}} + \text{MAE}_{\text{SSIM}}\right)$$

$$\text{MAE}_{\text{LPIPS}} = \frac{1}{N}\sum_{l=1}^{N}|\widehat{l_i} - l_i| \qquad \text{MAE}_{\text{PSNR}} = \frac{1}{N}\sum_{i=1}^{N}|\widehat{p_i} - p_i| \qquad \text{MAE}_{\text{SSIM}} = \frac{1}{N}\sum_{i=1}^{N}|\widehat{s_i} - s_i|$$

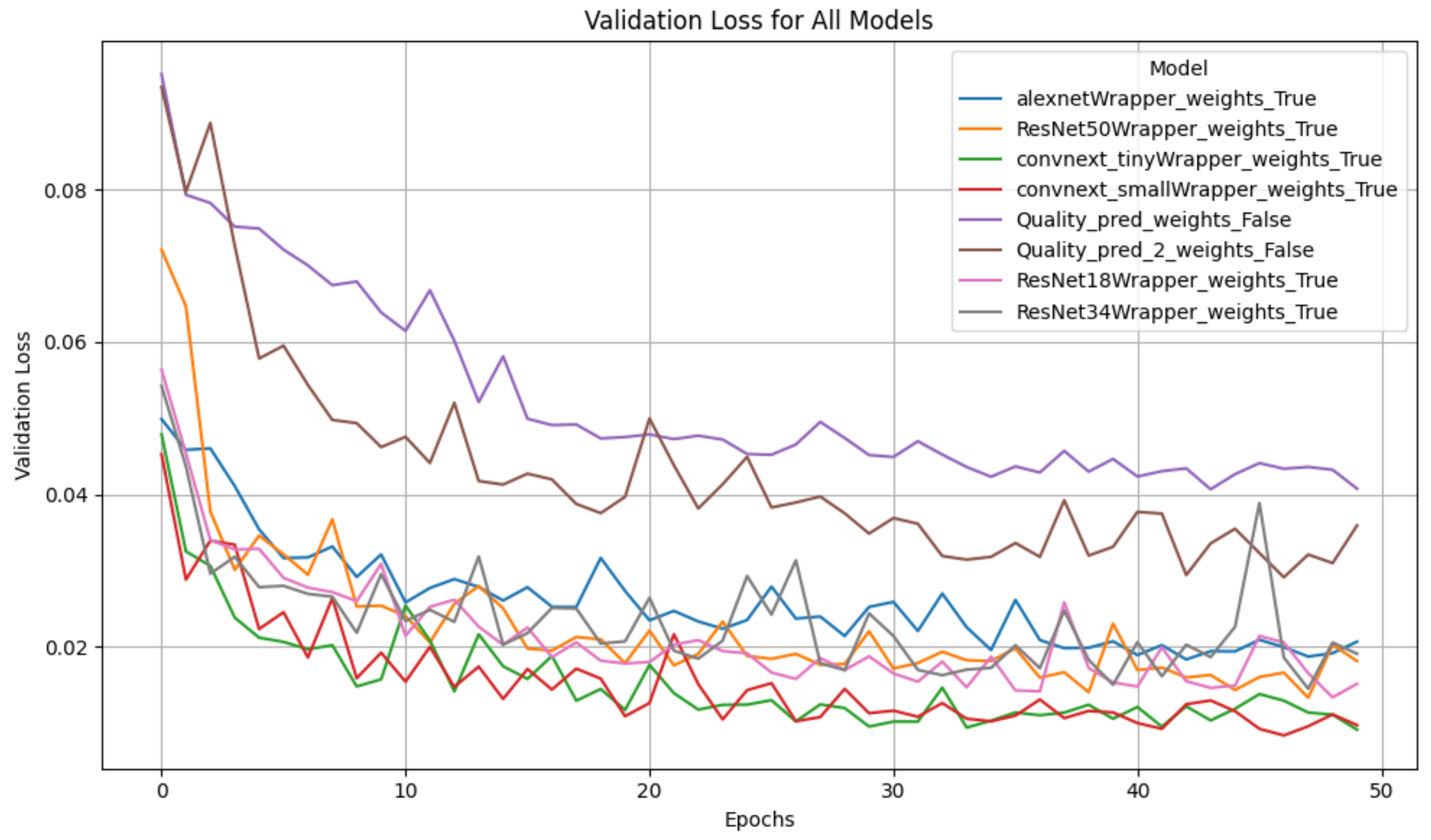

***Figure 2:** Validation loss over epochs for each model, used to identify the best-performing architecture for this task.*

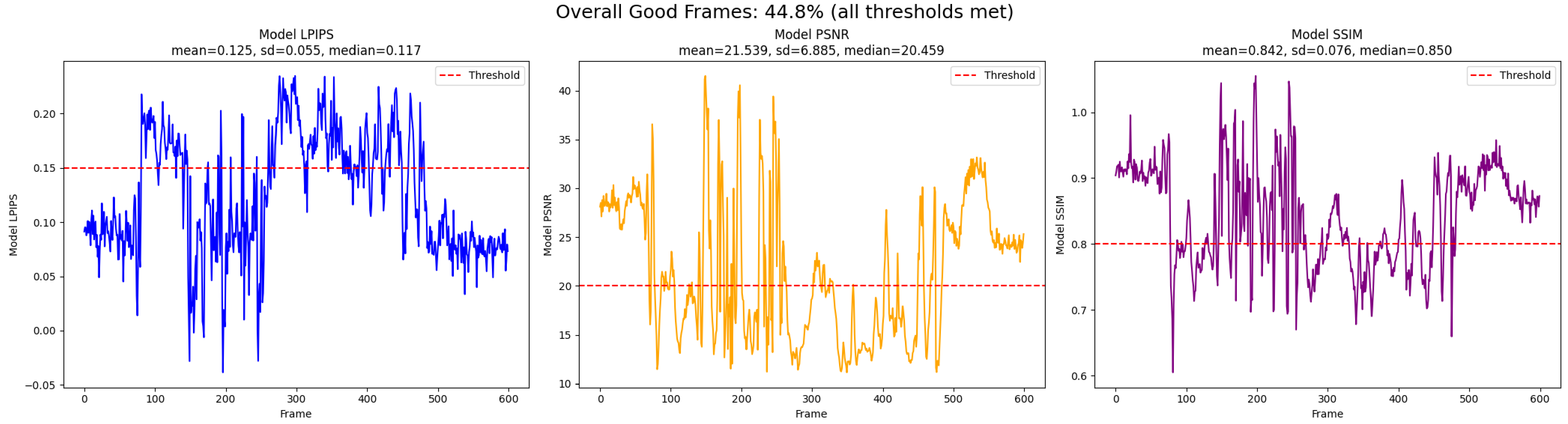

***Figure 4:** NoVisIQ applied to a real-world cholecystectomy surgery video, providing comprehensive video analysis and valuable information for medical professionals. When applied to real-world videos, NoVisIQ predicts all three metrics at 0.1 seconds per frame. With user-defined thresholds, it outputs the mean, median, and standard deviation for each metric, as well as the percentage of "good" frames based on those thresholds.*

## RESULTS

After training, we selected the ConvNeXt Small architecture [5] as the most suitable model for this task. When evaluated on synthetically augmented, previously unseen videos, the model closely tracked the true full-reference values, achieving an average distance of 0.0349 for LPIPS, 0.8547 for PSNR, and 0.0060 for SSIM between the model predictions and the true metrics.

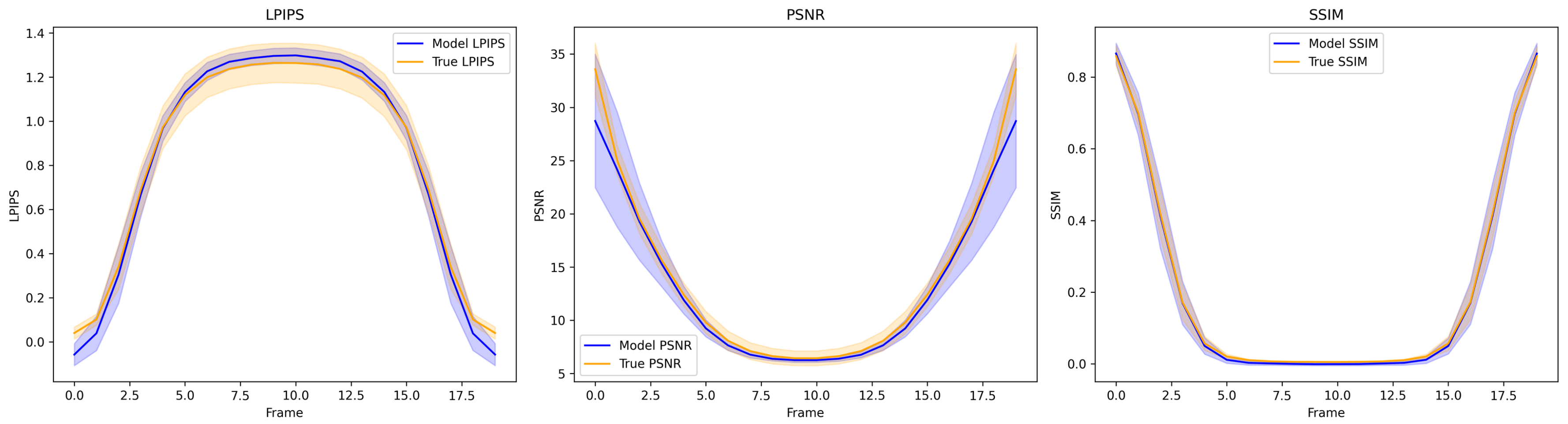

***Figure 3:** Model predictions closely match true LPIPS, PSNR, and SSIM values across frames of unseen, degraded videos, demonstrating accurate no-reference quality estimation.*

## CONCLUSIONS

NoVisIQ offers a robust solution for assessing video frame quality by accurately predicting three full-reference metrics without requiring access to original images at inference time. This supports object detection systems in identifying and managing low-quality frames, providing context to compromised performance and supporting downstream processing.

### Limitations:

NoVisIQ currently performs best on synthetic distortions similar to those seen during training, and there remains a gap in generalization to real-world video distortions. Additionally, the current model processes each frame independently, without leveraging temporal context from the surrounding frames, which limits its ability to handle severely obstructed or ambiguous frames.

### Next Steps:

To address these limitations, we plan to develop a streaming video model that incorporates temporal context. By processing sequences of frames, this approach will enable the model to better understand and assess obstructed or low-quality frames, improving robustness and overall video quality assessment in real-world scenarios.

## References

**[1]** Zhang, R. et al. (2018). *The Unreasonable Effectiveness of Deep Features as a Perceptual Metric*. arXiv preprint arXiv:1801.03924. https://arxiv.org/abs/1801.03924

**[2]** *TorchMetrics — Peak Signal-to-Noise Ratio (PSNR).* Lightning AI. https://lightning.ai/docs/torchmetrics/stable/image/peak_signal_noise_ratio.html

**[3]** Wang, Z., Bovik, A. C., Sheikh, H. R., & Simoncelli, E. P. (2004). *Image quality assessment: From error visibility to structural similarity.* IEEE Transactions on Image Processing, 13(4), 600–612. https://ece.uwaterloo.ca/~z70wang/publications/ssim.html

**[4]** *Natural Scene Statistics-Based Blind Image Quality Assessment.* Laboratory for Image and Video Engineering, The University of Texas at Austin. https://live.ece.utexas.edu/research/Quality/nrqa.htm

**[5]** Liu, Zhuang, et al. "A ConvNet for the 2020s." arXiv preprint arXiv:2201.03545 (2022). https://arxiv.org/abs/2201.03545

Lawrence Livermore National Laboratory
LLNL-POST-2009104
Prepared by LLNL under Contract DE-AC52-07NA27344.