Proceedings of the 6th International Conference on Fluid Flow and Thermal Science (ICFFTS 2025)
Barcelona, Spain - October 29 - 31, 2025
Paper No. 146
DOI: 10.11159/icffts25.146

Development and Evaluation of an 8x Frame Interpolation Method Based On Deep Learning for PIV Measurement

Erina Yamakawa[1], Toshiyuki Asano[2], Naoyuki Aikawa[1]
[1]Tokyo University of Science/Graduate School of Advanced Engineering
6-3-1 Niijuku, Katsushika-ku, Tokyo, Japan
8125537@ed.tus.ac.jp; ain@rs.tus.ac.jp
[2]NIIGATA Co.,Ltd.
2-12-5 Komaoka, Tsurumi-ku, Yokohama, Kanagawa
asano@ni-gata.co.jp

Abstract –Particle Image Velocimetry (PIV) is a widely used measurement method for visualizing velocity fields in fluids, and the use of high frame rate cameras is essential to ensure high temporal resolution. However, the high cost of introducing such cameras remains a major challenge. In this study, we propose a method for achieving high frame rate images by frame interpolation using deep learning on images acquired with general low frame rate cameras. Specifically, based on the FLAVR framework, we designed a new data structure and learning settings optimized for PIV applications, enabling 8x time interpolation. The interpolation results are evaluated using quantitative metrics (PSNR, SSIM) and qualitative particle tracking evaluation, demonstrating the effectiveness of our method through comparison with conventional methods.

***Keywords*:** Particle Image Velocimetry, frame rate, frame interpolation, deep learning, 3DCNN

1. Introduction

In recent years, PIV (Particle Image Velocimetry) has been widely used in various fields such as aviation, automotive, and food industries as a non-contact and visible method for quantifying fluid motion characteristics [1]. In particular, with the improvement of image-related equipment and computer performance, the spatial and temporal resolution of PIV has improved dramatically, enabling more precise fluid analysis.

In PIV, the movement of particles in a fluid is captured using a high-speed camera, and the velocity field is visualized by analyzing the movement of particles between consecutive images [2]. However, the accuracy of particle motion visualization depends heavily on time resolution. Therefore, it is generally considered essential to use a high-speed camera capable of capturing hundreds to thousands of frames per second (fps), but the cost of introducing such a camera is high, and its installation and operation require advanced specialized knowledge [3],[4]. Such technical and economic constraints are one factor limiting the application range of PIV.

To address this issue, frame interpolation techniques have emerged as a promising approach. These methods generate intermediate frames between two consecutive frames, thereby artificially enhancing temporal resolution. Among existing approaches, interpolation techniques can largely be classified into two categories: optical flow-based methods and deep learning-based end-to-end models.

Optical flow-based methods estimate explicit pixel-wise correspondences between frames to synthesize the intermediate frame. Representative examples include EQVI (Enhanced Quadratic Video Interpolation) [5], which demonstrate high performance in scenarios involving large motion or motion blur. However, their accuracy often degrades in PIV images due to the presence of fine-grained, stochastic particle patterns, where optical flow estimation becomes unreliable. In contrast, end-to-end deep learning methods directly generate intermediate frames from multiple inputs, without relying on explicit motion estimation. One such method is FLAVR (Flow-Agnostic Video Representations for Fast Frame Interpolation) [6], which leverages 3D convolution to jointly learn spatial and temporal features, achieving high interpolation quality across various scenes.

Nonetheless, most of these methods are designed for natural scenes or video compression contexts, and are not optimized for the nonlinear, locally complex particle motions observed in PIV data. In particular, maintaining particle trajectory consistency and density is critical for PIV applications, and visual smoothness alone is insufficient.

In this study, we propose an interpolation framework tailored for PIV, utilizing the FLAVR [6] architecture as a backbone. While preserving the original network structure, we customize the dataset and loss functions to accommodate the characteristics of particle images. Our model enables the generation of seven intermediate frames from low frame rate input (60fps) to simulate high frame rate output (480fps) in a single inference pass. Furthermore, we evaluate the proposed method in comparison with a conventional optical flow-based approach (EQVI [5]) in terms of particle trajectory accuracy, temporal resolution, and quantitative metrics such as PSNR and SSIM. The results demonstrate the effectiveness of our method and its potential to serve as a cost-efficient alternative to high-speed cameras in PIV applications.

2. Proposed Method for PIV Frame Interpolation

In this study, we propose a method for interpolating intermediate frames equivalent to high frame rates from particle images captured at low frame rates, with the aim of improving the temporal resolution of PIV. Specifically, we adopt a model based on a 3D convolutional neural network (3DCNN) that does not depend on optical flow [7]. Using this model, we examine methods for interpolating multiple frames between the first frame and the next frame. The proposed method accurately estimates the movement of each pixel between consecutive frames and synthesizes intermediate frames based on that movement information to generate visually smooth images.

2.1. Model Architecture

The model based on a 3D convolutional neural network (3DCNN) that does not depend on optical flow has a symmetric U-Net-style structure consisting of an encoder and a decoder, and the architecture is as shown in Fig. 1. Additionally, the model takes nine consecutive images (e.g., im1 to im9) as input and outputs seven intermediate frames between them in a single batch (equivalent to 8x interpolation). Compared to the conventional method of "sequentially interpolating one frame at a time," this approach is expected to prevent the accumulation of interpolation errors while maintaining the consistency of particle trajectories.

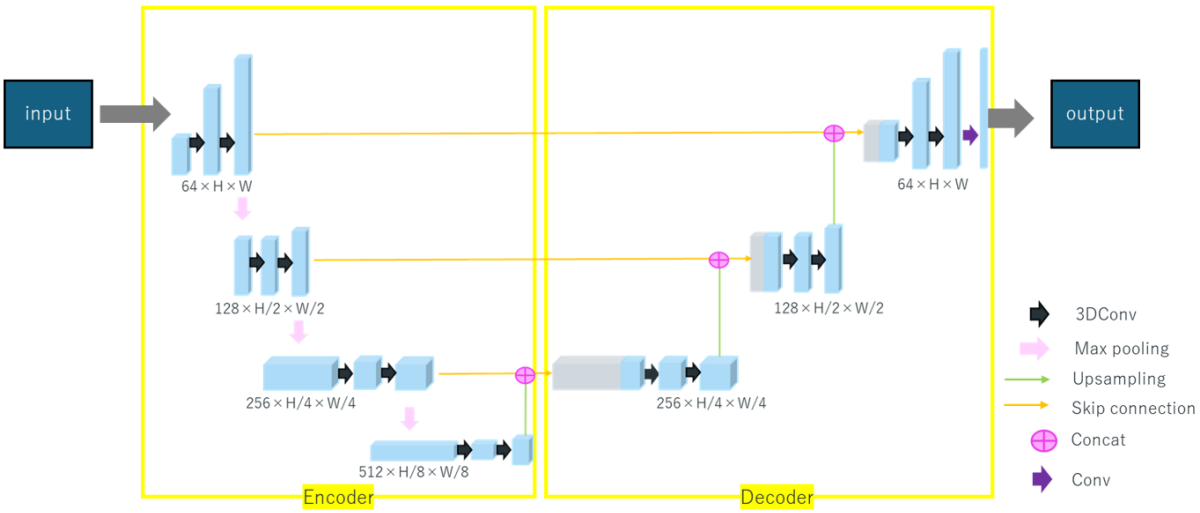


Fig. 1: Model for frame interpolation for PIV.

As shown in Fig. 1, a 3D convolution kernel is used for the main operations in the encoder and decoder. The 3D kernel, which has time dimension t and spatial dimensions H and W, can simultaneously capture temporal motion and changes between frames and spatial information within frames.

The encoder applies 3D convolution, including time and space directions, to effectively extract spatiotemporal features such as motion and changes between multiple frames, background consistency, and continuous shape changes of animal bodies. The extracted features are accumulated in multiple layers while reducing resolution, simultaneously capturing local and global information in time and space. The output of each stage of the encoder is sent to the corresponding stage of the decoder via skip connections (orange arrows). This is an important feature of U-Net, enabling the generation of more detailed outputs by supplying high-resolution spatial information to the decoder. On the other hand, the decoder restores the resolution through upsampling while retaining spatial detail information through skip connections from the encoder. This improves the accuracy of reconstructing minute particle movements.

2.2. 3D convolution block

The 3D convolution blocks used in each layer (see Fig. 2) enable feature extraction that takes into account not only temporal and spatial continuity but also correlations and dependencies between channels.

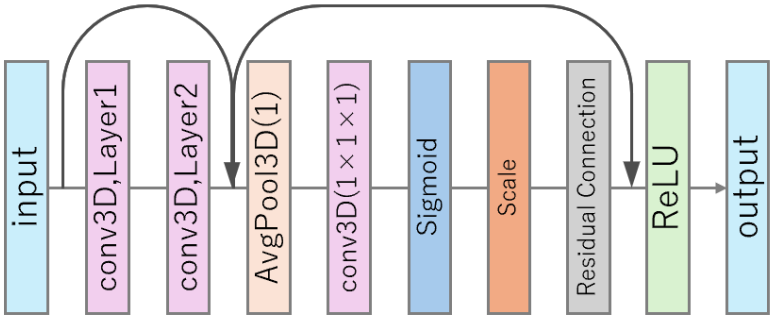


Fig. 2: 3D convolution block.

This allows us to accurately capture local micro-changes in particles and contrast differences with the background, resulting in a stable feature map that is resistant to noise. In addition, some blocks undergo reweighting similar to the channel attention mechanism, enabling selective focus on areas with movement.

2.3. Optimization

The Adam optimizer was used for learning, and the learning rate was adjusted using the ReduceLROnPlateau scheduler. This allowed for dynamic adjustment to suppress overfitting while achieving stable convergence [8]. As the loss function, L_1 loss was adopted to minimize the difference between the reconstructed image and the correct image.

$$L_1 = \frac{1}{N}\sum_{i,j=1}^{N}\left|\hat{y_{(i,j)}} - y_{(i,j)}\right| \tag{1}$$

where N is the total number of pixels, $\hat{y_{(i,j)}}$ is the predicted value, and $y_{(i,j)}$ is the correct value. The difference between the predicted image and the original image is measured by summing the absolute error for each pixel. In addition, the weights of the model that showed the best performance during training were saved and configured so that they could be used for resuming training and testing evaluation.

3. Experiment

3.1. Experimental Overview and Conditions

We designed an observation flow channel with a rectangular obstacle placed in the center of the flow channel, as shown in Fig. 3. Water supply and drainage tanks were installed at both ends of the flow channel to generate a stable steady flow.

For flow condition imaging, we combined a high-speed camera HAS-D73 (Detect Co., Ltd.) with a sheet-shaped laser illumination system and captured images at a resolution of 1280 × 1028 pixels, 480 fps.

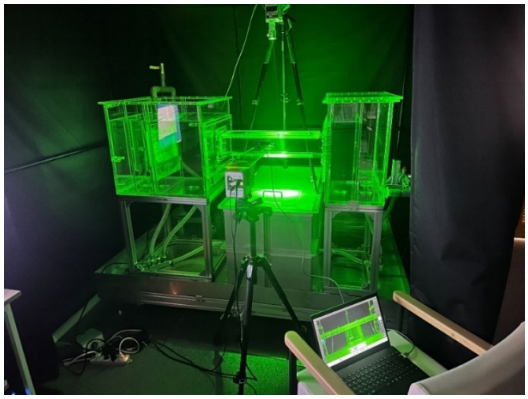


Fig. 3: Experimental setup.

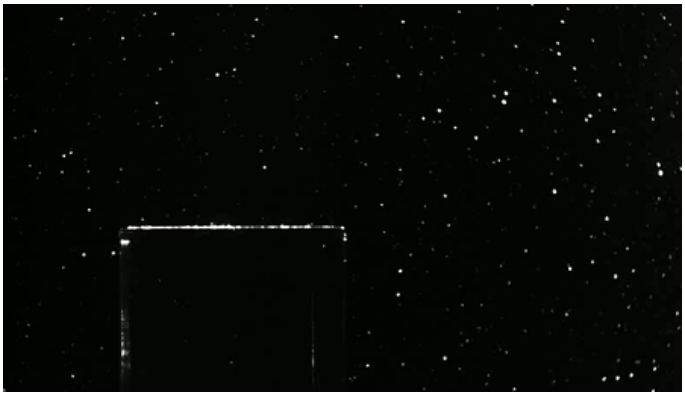


Fig. 4: PIV measurement image with 448 × 256 pixels.

To focus on the flow around the obstacles, a region of interest (ROI) of 448 × 256 pixels was extracted from the imaging frame (1280 × 1028 pixels) (see Fig. 4). Subsequent analysis and frame interpolation were performed on this ROI.

To verify the effectiveness of the proposed method, low-frame-rate images (60 fps) were generated by downsampling high-frame-rate images (480 fps) by a factor of 8. Four high-frame-rate images (480 fps) captured over 1 minute and 28 seconds were used as training data. Test data was obtained from 15 seconds of low-frame-rate images not included in the training data.

The detailed training conditions are summarized in Table 1.

Table 1: Learning conditions.

Item	Details
Training samples/sec	88
Test samples/sec	15
Image Size/pixel	448×256
Optimizer	Adam
Loss	L_1
Batch Size	16
Epoch	200
Learning Rate	2×10^{-4}

3.2. Experimental Results

To evaluate the accuracy of frame interpolation from a low to a high frame rate, we performed an objective assessment using two widely adopted image-quality metrics: Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index Measure (SSIM). Both metrics quantify the similarity between the interpolated and ground-truth images and are regarded as reliable indicators of perceptual quality.

PSNR expresses, on a logarithmic scale, the ratio between the square of the maximum possible pixel value and the mean-squared error (MSE) between two images, and is defined as follows [9]:

$$PSNR = 10log_{10}\left(\frac{255^2}{MSE}\right) \tag{2}$$

where 255 is the maximum pixel value in an 8-bit image, and MSE denotes the mean square error between corresponding pixels of the interpolated and reference images. A larger PSNR indicates lower reconstruction error and correspondingly higher fidelity.

In contrast, SSIM measures perceptual similarity by jointly evaluating luminance, contrast, and structural information; it is defined as [9]:

$$SSIM = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)}. \tag{3}$$

Here μ_x and μ_y are the average brightness of the interpolated image and the correct image, respectively; σ_x and σ_y are their standard deviations and σ_{xy} is the covariance between them. The constants c_1 and c_2 are small positive numbers introduced to prevent division by zero. SSIM ranges from 0 to 1, with values closer to 1 indicating higher structural similarity.

Table 2 shows the average PSNR and SSIM values calculated for the results of interpolating low frame rate images of 60 fps to 480 fps using conventional methods and the proposed method.

Table 2: Average evaluation index of frame interpolation.

Method-based	PSNR[dB]	SSIM
EQVI [5]	44.5289	0.9924
Proposed method	49.8426	0.9870

Table 2 reports that the proposed method attains a PSNR of 49.8426 dB, outperforming EQVI by 5.3 dB, respectively. On the other hand, SSIM decreased slightly, which is thought to be due to the influence of fine structures and blurred contours unique to particle images. In particular, since SSIM is sensitive to structural consistency and local sharpness, the proposed method's smooth and continuity-oriented interpolation process may have slightly decreased structural similarity in some areas.

To evaluate the quality of frame interpolation, we first present the ground-truth trajectories. Fig. 5 shows tracer-particle paths captured at a high frame rate (480 fps) over nine consecutive frames. Motion between frames is color-coded chronologically in eight hues: blue, green, red, cyan, yellow, magenta, gray, and purple.

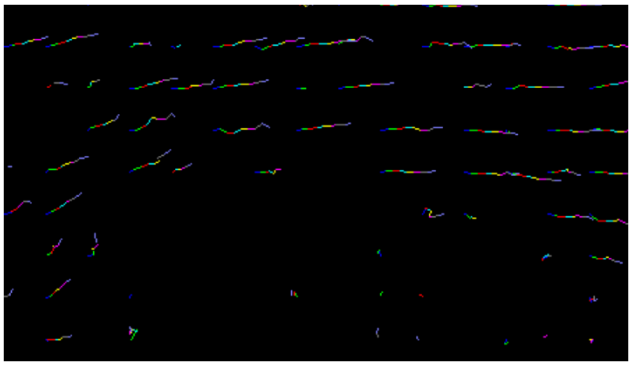


Fig. 5: Trajectory diagram using high frame rate frames.

Figure 6 uses the same color scheme to display trajectories obtained with the conventional technique, Extended Quadrature Video Interpolation (EQVI) [5], while Fig. 7 presents the results produced by the proposed method.

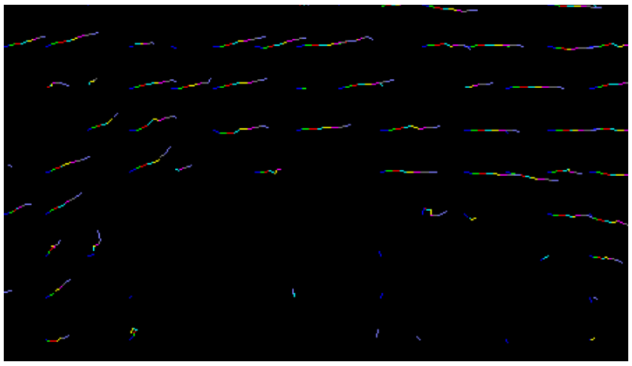


Fig. 6: Trajectory diagram using conventional method.

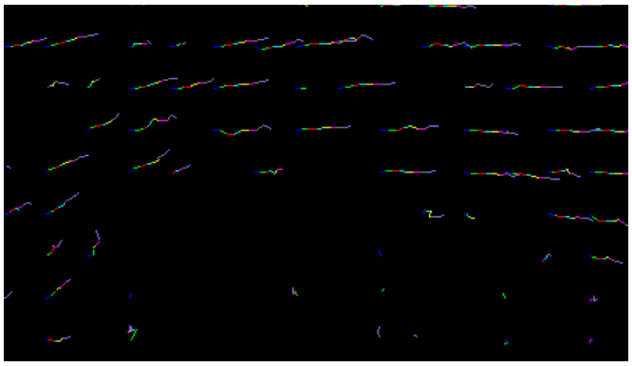
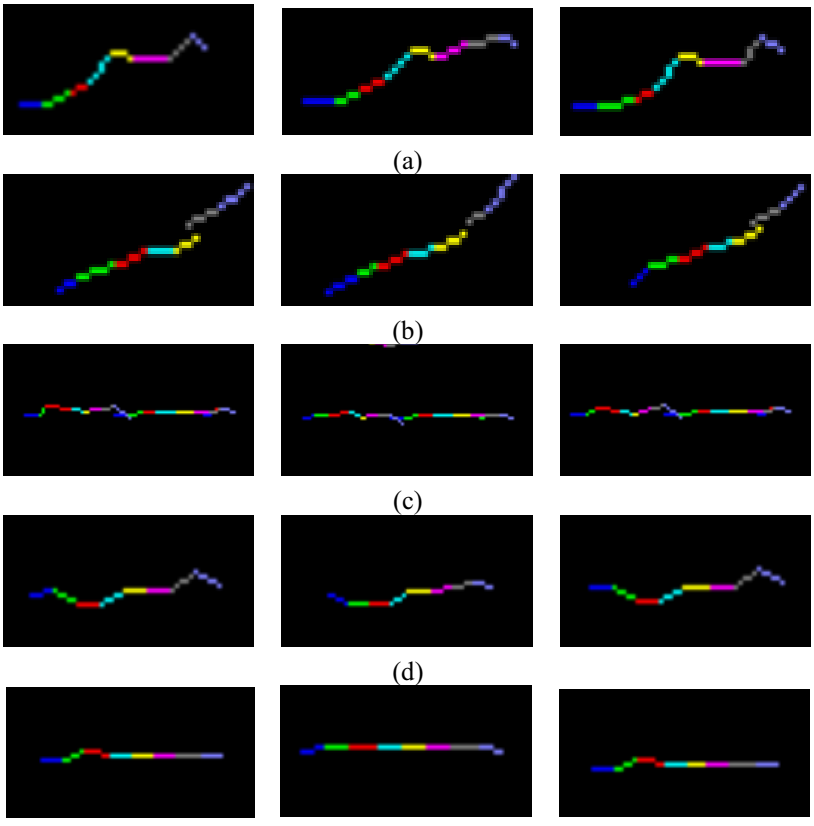


Fig. 7: Trajectory diagram based on proposed method.

Compared to the ground truth in Fig. 5, EQVI reproduces the overall motion, but renders several paths excessively linear. Noticeable discrepancies appear in path direction and color transitions, indicating that the conventional model does not adequately capture nonlinear motion. In contrast, the proposed method generates smoother, more curved trajectories and preserves color gradients that correspond to local velocity variations, demonstrating superior fidelity.

To examine local accuracy in greater detail, Fig. 8 enlarges five representative regions (a)–(e).



(a)

(b)

(c)

(d)

(e)

Fig. 8: Trajectory diagram of a certain part (from left to right: correct image, interpolation result using the conventional method, interpolation result using the proposed method).

For each region, the left panel shows the ground truth (Fig. 5), the center panel shows the EQVI results (Fig. 6), and the right panel shows the results of the proposed method (Fig. 7). In the ground-truth panels, curved trajectories and subtle speed changes are expressed as continuous color gradients. Although EQVI retains some nonlinearity, its trajectories often deviate from the ground truth and appear monotonous, especially when paths curve gently or velocity changes abruptly. The proposed method reproduces curvature and directional changes more faithfully, and its color distributions closely match those of the ground truth. However, slight discrepancies remain in high-speed regions (e.g., the yellow areas in panels (b) and (e)), suggesting the need for further refinement of local velocity estimation in future work.

4. Conclusion

In this study, we proposed a method for converting low frame rate images to high frame rate images using deep learning to address the issues of high cost and difficulty associated with high-speed cameras for PIV applications. Specifically, we extended the FLAVR [6] model based on a 3D convolutional neural network (3D-CNN) [7] and constructed an 8x interpolation model that generates seven intermediate frames equivalent to 480 fps from 60 fps input. By combining a dataset and loss function specialized for particle images, we simultaneously learned temporal and spatial features to improve the reproducibility of nonlinear particle trajectories. As a result, we demonstrated higher interpolation performance than conventional methods in quantitative evaluation (PSNR), and confirmed that we could more realistically reproduce complex particle motion in comparisons using trajectory diagrams. On the other hand, the slight decrease in SSIM is thought to be due to the fine structure and blurred contours characteristic of particle images. In the future, we aim to improve the practicality of fluid analysis by introducing interpolation methods that further improve the accuracy of responding to nonlinear motion and optimizing the model structure that takes into account the spatial distribution of tracer particles.

References

[1] S. Cai, C. Gray and G. E. Karniadakis , “Physics-Informed Neural Networks Enhanced Particle Tracking Velocimetry: An Example for Turbulent Jet Flow,” *IEEE Trans.. on Instrum. Meas.*, vol. 73, 2024.

[2] M. Tanahashi, Y. Fukuchi, Choi Gyung Min, Katsuhiko Fukuzato and Toshio Miyauchi, “High Spatial and Temporal Resolution Time-Series Particle Image Velocimetry,” *Proc. JSME*, vol. 81, p. 145, 2003.

[3] T. Ishima, “Fundamentals of Partical Image Velocimetry (PIV),” *J. Combust. Soc. Japan*, 66(197), pp. 224-230, 2019.

[4] M. Ishikawa, K. Okamoto and H. Masarame, “Development of the Multi-Scale PIV System Using High-Speed Cameras,” *Visualization Soc. Japan*, 25(10), pp. 64-71, 2005.

[5] Y. Liu, L. Xie, L. Siyao, W. Sun, Y. Qiao and C. Dong, “Enhanced Quadratic Video Interpolation,” *ECCV Workshops*, pp. 41-56, 2021.

[6] T. Kalluri, D. Pathak, M. Chandraker and D. Tran, “FLAVR: Flow-Agnostic Video Representations for Fast Frame Interpolation,” *WACV*, 2023.

[7] S. Ji, W. Xu, M. Yang and K. Yu, “3D Convolutional Neural Networks for Human Action Recognition,” *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 35, issue. 1, pp. 221-231, 2012.

[8] A. Sunyoto, Y. Pristyanto and A. F. Nugraha, “Enhanced Classification of Potato Leaf Disease Using Xception and ReduceLROnPlateau Callbacks,” *Proc. IEEE Int. Conf. Internet of Things and Intelligence Systems (IoTaIS)*, Bali, Indonesia, 2023.

[9] A. Horé and D. Ziou, “Image Quality Metrics: PSNR vs. SSIM,” *Proc. ICPR*, 2010.