A Forensic Video Upscaling Colorizing and Denoising Framework for Crime Scene Investigation

S. Prema and S. Anita

Abstract Digital videos have been widely used as key evidence sources in Forensic crime scene investigations. Resolution is one of the most dominating parameter which affects the overall quality of the video. The main goal of this paper is to find an efficient forensic video analysis framework to assist the forensic crime scene investigation. A forensic video analysis framework (FVAF) that employs an efficient video enhancing deep learning model for increasing resolution of the low quality videos is used. The low resolution video is fed as input to the model. First, the video is pre-processed using fastai deep learning library. Large videos are cropped to manage runtime efficiently. Second, the video is rescaled for increasing the resolution by Spatial Resolution method. The framework successfully increases the resolution of the video from SD-standard definition Resolution type of 480p with Aspect Ratio 4:3 of Pixel size 640 × 480 to Full Ultra HD Resolution type of 8K or 4320p with Aspect Ratio 16:9 of Pixel Size 7680 × 4320. The rescaled videos are submitted for colorization process. DeOldify deep learning model using Self-Attention Generative Adversarial Network and Two Time-Scale Update Rule is adopted by FVAF framework for colorizing the videos. Also, the colorized videos are trained and tested by various video enhance AI models model Gaia High Quality 4K rendering and Theia fine Tune detail. 4K not rendered and Theia Fine Tune Fidelity: 4K not rendered and video denoise AI models model Standard, clear, lowlight, severe noise and Raw. The upscaled and colorized video is also trained and tested using denoise video enhance AI and video denoise AI models. The results of each model are stored for comparison. From the stored results best video enhance AI model and the best video denoise AI models is selected. Lowlight AI model and Gaia high quality 4K rendering are used

---

Electronic supplementary material The online version of this chapter (http://doi.org/10.1007/978-981-99-5091-1_18) contains supplementary material, which is available to authorized users.

S. Prema
Arulmigu Arthanareeswarar Arts and Science College, Thiruchengodu, Tamil Nadu, India

S. Anita (✉)
Bhusanayana Mukundadas Sreenivasaiah College for Women, Bengaluru, Karnataka, India
e-mail: anita.sigamani@gmail.com

S. J. Patel et al. (eds.), *Information Security, Privacy and Digital Forensics*,
Lecture Notes in Electrical Engineering 1075,
https://doi.org/10.1007/978-981-99-5091-1_18

in this FVAF to produce high standard video for Forensic Analysis. We run this model using GPU to efficiently pre-process the video. By this framework, we increase the resolution of the video footages to further assist the forensic crime investigation.

Keywords Forensic · Video enhancement · Spatial resolution · Colorization · Deep learning

1 Introduction

Forensic Video Analysis plays an important role in the evaluation of video. It is often necessary for a forensic technician, analyst, or video forensic expert to perform a digital video evidence recovery in order to secure the Digital Media Evidence and establish a chain of custody. LEVA defined Digital Media Evidence as "Information of probative value stored in binary form" [1]. CCTV surveillance video recordings are the most common type of digital media evidence (DME). Once the recordings have been secured, an accurate chain of custody can be presented to the trier of fact. In addition, the forensic expert that acquired the video evidence can ensure that the highest quality versions of the recording are obtained so that a successful forensic video enhancement or forensic image comparison can be performed. The objective of Forensic Video Enhancement is to clarify or enhance the events as they occurred. This is done using nondestructive techniques to preserve the video evidence integrity, and pixel quality. Clarifying or enhancing the events as they occurred assists the Trier of Fact to weigh the video evidence and its relevance to the litigation. Video forensic experts are often asked to enhance CCTV Surveillance video recordings and to provide video image enhancement for identification purposes in the court. Popular enhancement techniques are applied to an investigation are Scaling [2], Pixel Interpolation, sharpening, warp stabilization, Shadow and highlight adjustments, frame averaging, pixel aspect ratio calibration, Color Correction, Reverse Projection, Photogrammetry, Motion Tracking, Demonstrative Video Exhibits. Digital videos are used as key evidence sources in Forensic crime scene investigations. Resolution is one of the most dominating parameter which affects the overall quality of the video. The objective of this paper is to develop Advanced Forensic Video Restoration Framework to perform accurate investigation.

1.1 Forensics

Forensic science, also known as criminalistics, is the application of science to criminal and civil laws, mainly on the criminal side during criminal investigation, as governed by the legal standards of admissible evidence and criminal procedure. Forensic science is a broad field that includes; DNA analysis, fingerprint analysis, blood stain pattern analysis, firearms examination and ballistics, tool mark analysis, serology, toxicology, hair and fiber analysis, entomology, questioned documents, anthropology, odontology, pathology, epidemiology, footwear and tire tread analysis, drug chemistry, paint and glass analysis, digital audio video, and photo analysis. Forensic scientists collect, preserve, and analyze scientific evidence during the course of an investigation. While some forensic scientists travel to the scene of the crime to collect the evidence themselves, others occupy a laboratory role, performing analysis on objects brought to them by other individuals. Still, others are involved in the analysis of financial, banking, or other numerical data for use in financial crime investigation, and can be employed as consultants from private firms, academia, or as government employees. Computers are used for committing crimes, and, thanks to the burgeoning science of digital evidence forensics, law enforcement now uses computers to fight crime. Digital evidence is information stored or transmitted in binary form that may be relied on in court. It can be found on a computer hard drive, a mobile phone, among other places. Forensic video analysis is the scientific examination, comparison, and/or evaluation of video in legal matters. The video forensic process must be performed in a forensic lab that is equipped with the appropriate tools and follows best practice protocols in order to process the video recording with integrity and accuracy.

2 Literature Review

2.1 Survey on Deep Learning Algorithms

Deep learning has evolved over the past five years, and deep learning algorithms have become widely popular in many industries. It is a type of machine learning that works based on the structure and function of the human brain. Industries such as health care, eCommerce, entertainment, and advertising commonly use deep learning (Table 1).

2.2 Architectures Design of Deep Learning Models

See Figs. 1, 2, 3, 4, and 5.

3 Video Super-Resolution Methods Based on Deep Learning: A Survey

See Table 2.

Table 1 Survey on deep learning algorithms

Deep learning algorithm	Developed by	Year	Components and operations	Output	Uses
Recurrent neural networks (RNNs) [3, 4]	David Rumelhart's	1986	• Training • Gradient descent • Global optimization method • Fully recurrent • Elman and Jordan networks	• (Batchsize, units)	• One vector per timestamp per sample
LeNet/ convolutional neural networks (CNNs) [5–7]	Yann LeCun	1988	• Multiple layers • Convolution layer • Pooling layer • ReLU activation function • Downsampling • Flattening	• Feature map • Linear vector	• Image processing and object detection • Identify satellite images
Long short term memory networks (LSTMs) [8, 9]	Juergen Schmidhuber's	1995	• Chain-like structure • Four interacting layers • Forget irrelevant parts • Update the cell-state values	• 4 different sets of results • Dault: last hidden state • Reurn_ sequences = true: all hidden states	• Speech recognition • Music composition • Pharmaceutical development
Generative adversarial networks (GANs) [10–12]	Ian Goodfellow	2014	• A generator • A discriminator • Initial training • A zero-sum game	• Inception score • Inception-v3 • Fréchet inception distance (FID)	• Improve astronomical images

4 Problem Statement

Resolution is one of the most dominating parameter which affects the overall quality of the video. Resolution greatly influences the viewing experience. One of the key aspects of delivering a good quality video is understanding and determining the correct video resolution of an on-demand video or live stream. Comparing with the frame rate, Video quality feature is the most important to deal with. For performing effective analysis of crime cases in Forensic field, video enhancement plays a vital

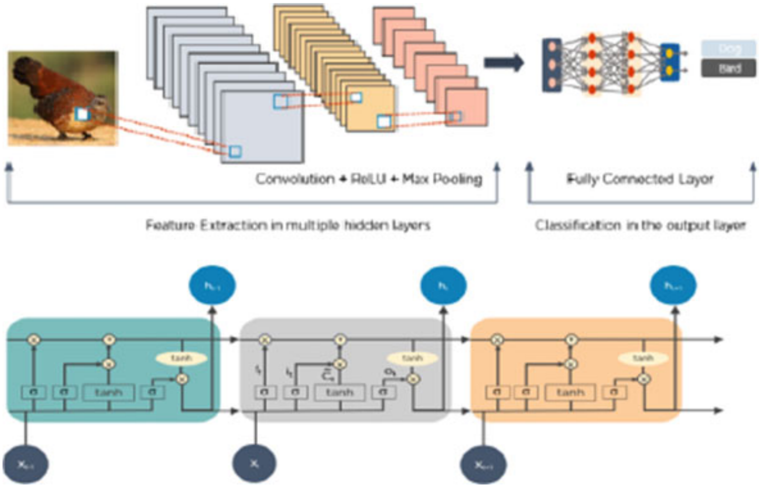

Fig. 1 CNN [7] and LSTMs flow of operation [8, 9]

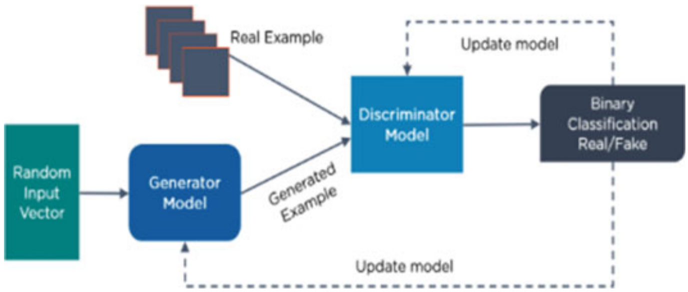

Fig. 2 Unfolded RNN [13] and a working process of GAN [12]

role. This paper is created with the objective of enhancing and colorizing the old footage videos for helping in efficient Forensic crime scene investigation.

5 Proposed Methodology

In this framework, we first pre-process the video. Towards pre-processing the video, we crop large videos to manage runtime efficiently. Next, we rescale the video for increasing the resolution. In our FVAF model, we use Spatial Resolution method for increasing the resolution of the low quality video. Our framework successfully increases the resolution of the video from SD-standard definition Resolution type of 480p with Aspect Ratio 4:3 of Pixel size 640 × 480 to Full Ultra HD Resolution type of 8K or 4320p with Aspect Ratio 16:9 of Pixel Size 7680 × 4320.

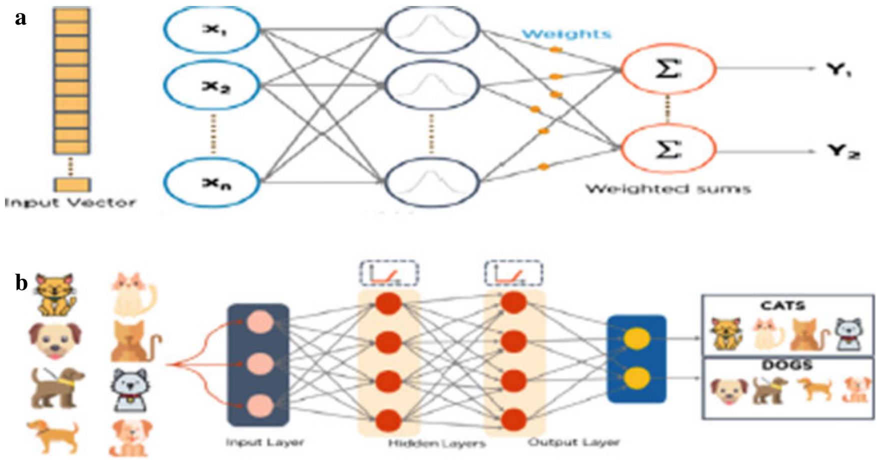

Fig. 3 **a** RBFN model [14] and **b** MLP process [15]

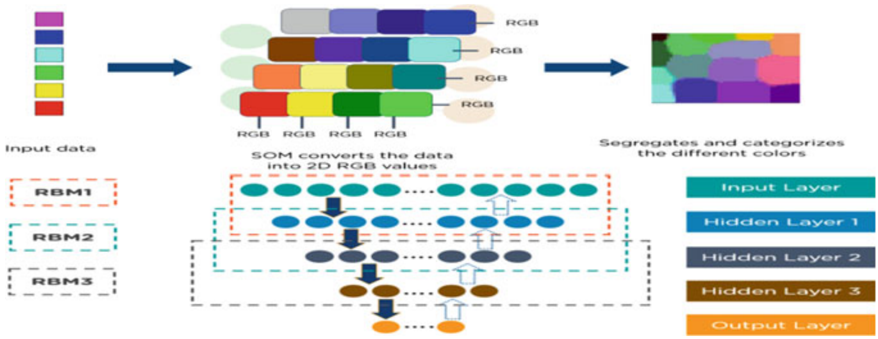

Fig. 4 SOMs process [16] and DBN architecture [17]

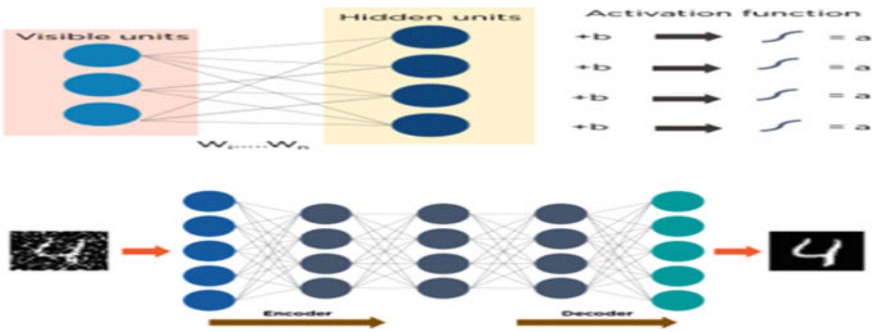

Fig. 5 RBMs function [18] and autoencoders operation flow [19]

Table 2 Existing video super-resolution methods based on deep learning

Model	Loss function	PSNR	SSIM	Ref.
Frame recurrent video super resolution (FRVSR) 2018	Loss with total variation regularization	26.17	0.798	[9]
Super-resolution optical flow for video super-resoltion (SOFVSR) 2019	MSE loss and MC loss	NA	NA	[18]
Motion estimation and motion compensation network (MEMC-Net) 2021	Charbonnier (Cb) loss	34.95	0.9679	[20]
Dual subnet and multistage communicated upsampling (DSMC) 2021a	Cb loss; perceptual loss; the dual loss	27.56	0.8934	[21]
Video restoration based on deep learning: comprehensive survey 2022	Reconstruction loss	NA	NA	[22]
Video super-resolution via dense non-local spatial–temporal convolutional network (DNSTNet) 2020	1-norm loss	NA	NA	[23]
Space-time-aware multi-resolution network (STARnet) 2020	Three losses	NA	NA	[1]
BasicVSR++ (VideoLDV dataset) 2022	Charbonnier loss [24]	31.63	NA	[9]
CDVSR (Video LDV dataset) 2022	Charbonnier loss [24]	23.03	NA	[21]

Also, this framework aims in colorizing black and white videos by using DeOldify deep learning model using Self-Attention Generative Adversarial Network and Two Time-Scale Update Rule. We run this model using GPU to efficiently pre-process the video. By this framework, we increase the resolution of the video footages to further assist the forensic crime investigation. The upscaled frame is shown in Fig. 6a and Colorized frame is shown in Fig. 6b. In our FVRF model we use Spatial Resolution method for increasing the resolution of the low quality video. Spatial resolution is the height and width of the frame, which is measured in pixels. It's the total number of pixels in every individual frame. DeOldify Model includes Self Attention GANs, Progressive Growing of GANs, and Two time scale update rule. In this proposed framework DeOldify deep learning model is used to colorize the video. The DeOldify model uses design specification of self-attention used in the "Self-attention Generative Adversarial Networks" [25]. "Traditional convolutional GANs generate high resolution details as a function of only spatially local points in lower resolution feature maps [25]. In SAGAN, details can be generated using cues from all feature locations [12]. The discriminator can check that highly detailed features in distant portions of the frames are consistent with each other." Deoldify using self-attention guarantees maximal continuity, consistency, and completeness in colorization [26]. GANs: Generative adversarial networks (GANs) are deep neural net architectures [24]. GANs are comprised of two nets namely Generator and Discriminator, trained

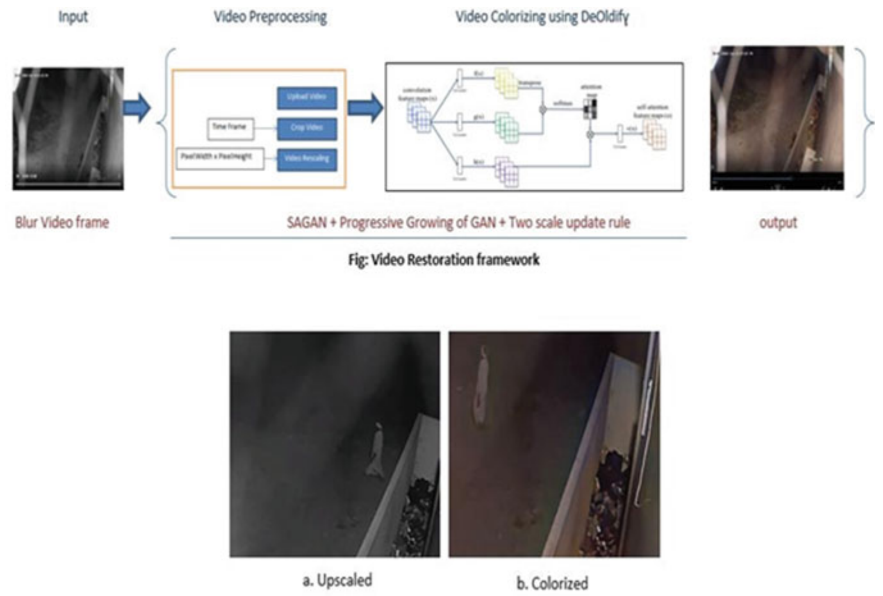

Fig. 6 Video restoration and colorization framework

one against the other. Typically, the generator is of main interest. The discriminator is an adaptive loss function that gets discarded once the generator has been trained. Generative model tries to fool the discriminator while discriminator acts as a detective trying to catch the forgers. In Progressive Growing of GANs the key takeaway is to grow both Critic and Generator model's layers progressively starting from a low resolution, and to add new layers that model increasingly fine details as the training progresses [27]. Two time scale update rule: It is choosing different learning rates for the generator and discriminator. DCGAN [10] and WGAN-GP [28] using different learning rates have achieved state-of-the-art results. 'Using different learning rates'—Is choosing a higher learning rate for the discriminator and a lower one for the generator. ConvBlock is used to build the Unet Block which in turn is used to build the Unet Model [29]. U-Net based generators are used in DeOldify. The Unet block uses Conv2D, activation layer, and Batch normalization modules.

5.1 Notation

The notations used are: x: Real data, z: Latent vector, G(z): Fake data, D(x): Discriminator's evaluation of real data, D(G(z)): Discriminator's evaluation of fake data, Error(a,b): Error between a and b [9].

5.2 The Discriminator

The goal of the discriminator is to correctly label generated images as false and empirical data points as true [9]. Therefore, we might consider the following to be the loss function of the discriminator:

$$LD = Error(D(x), 1) + Error(D(G(z)), 0) \tag{1}$$

5.3 The Generator

We can go ahead and do the same for the generator. The goal of the generator is to confuse the discriminator as much as possible such that it mislabels generated images as being true.

$$LG = Error(D(G(z)), 1) \tag{2}$$

5.4 Binary Cross Entropy

A common loss function that is used in binary classification problems is binary cross entropy. As a quick review, let's remind ourselves of what the formula for cross entropy looks like:

$$H(p, q) = Ex \sim p(x)[-log q(x)] \tag{3}$$

5.5 Training the Discriminator

When training a GAN, we typically train one model at a time. In other words, when training the discriminator, the generator is assumed as fixed. In min–max, the quantity of interest can be defined as a function of GG and DD. Let's call this the value function:

$$V(G, D) = Ex \sim p_{data}[log(D(x))] +_{Ez \sim pz}[log(1 - D(G(z)))] \tag{4}$$

6 Results

6.1 Video Upscaling and Sharpening

The upscaled and colorized video is further trained with AI models [11] for producing better results. The video is upscaled with theia models in theia-Detail + theia fine tune-Fidelity, Artemis Aliased, and Moire and Gaia models in Gaia High Quality 4K rendering. A more enhanced clarity of the sky and the vehicles moving on the road are produced using theia fine tune-Fidelity: 4K. The results obtained are as follows (Figs. 7, 8, and 9).

Fig. 7 **a** Original scaled 50%, **b** Gaia high quality 4K rendering, **c** Theia fine tune fidelity: 4K not rendered, **d** Artemis aliased and Moire zoomed at 36%

Fig. 8 **a** Original scaled 50%, **b** Theia fine tune fidelity: 4K rendering, **c** Theia fine tune fidelity: 4K not rendered. **d** Artemis high quality: 4K not rendered, zoomed at 36%. Theia fine tune fidelity: 4K rendering is found to be best compared with Artemis high quality: 4K not rendered. The objects in Fig. 8b are seen with clear edges than in the output of Artemis high quality model

Fig. 9 Results produced by DAIN and Artemis Aliased and Moire

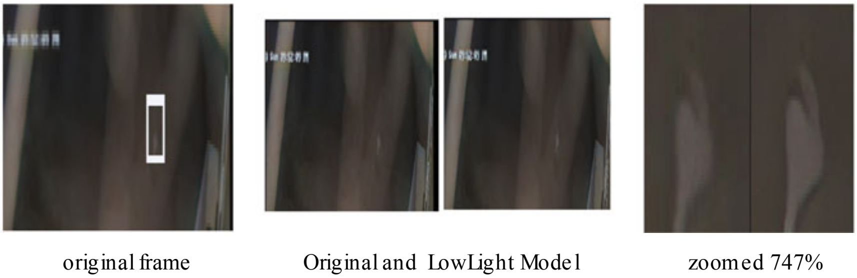

Fig. 10 The original video denoised with remove noise: 26 and enhance sharpness: 15. By setting removing noise to 26 and enhance sharpness to 15 and zooming at 747% the blur image has been sharpened and the noise level is reduced and the person wearing white shirt with pale red color ribbon on round his head is seen

6.2 Video Denoising

See Figs. 10 and 11.

The Artemis processed video is passed as input to the Video Enhance AI models Gaia High Quality 4K rendering, Theia fine Tune detail 4K not rendered, and Theia Fine Tune Fidelity: 4K not rendered at Zoom percent 59. Comparatively shows a clear view of the person walking with less artifacts.

6.3 Video Colorization

See Figs. 12, 13, 14, and 15.

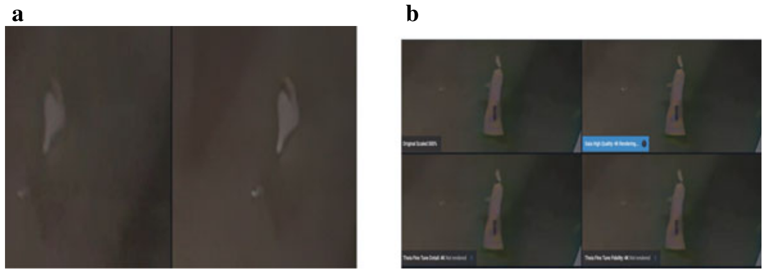

Fig. 11 The original video denoised with remove noise: 16 and enhance sharpness: 24. By setting removing noise to 16 and 3 enhance sharpness to 24 and zooming at 200% the person's shirt edges are seen clear and the image sharpness is increased. The video frame is extracted as an image and tested with denoise AI models standard, clear, lowlight, severe noise, and raw. Comparatively lowlight model has shown improved sharpness with reduced noise. **a** Original scaled 50%, **b** Gaia high quality 4K rendering, **c** Theia fine tune detail 4K not rendered, **d** Theia fine tune fidelity: 4K not rendered zoomed at 59%

Fig. 12 Video resolution: SD—standard definition resolution type of 480 pixels resolution with aspect ratio 4:3 of pixel size 640 × 480

Fig. 13 Video resolution: full ultra HD resolution type of 8K or 4320p with aspect ratio 16:9 of pixel size 7680 × 4320

Fig. 14 Upscaled video

Fig. 15 Video colorization using Deoldify

6.4 Future Research Direction and Challenges

1. **Combine Ultra-high resolution with high FPS:**
   High-speed imaging captures 1000 frames per second so that fast-moving subjects can be portrayed smoothly and fast motions can be analyzed with the constraint of lower resolution and several seconds of recording time.
2. **Extended Reality (XR): Immersive media and high-speed imaging:**
   Immersive media demands resolution above 8K. Technologies developed for "High-Speed Imaging" will be useful as more pixels are used.
3. **Handling distracting stutters and dropped frames is a challenge:**
   In video editing applications, slowing down footage from a 24 fps clip would result in distracting stutters and dropped frames. One can use chromos to create beautiful slow motion results regardless of the source frame rate.
4. **Maximizing FOV (Field of View) > 8K resolution is a challenge:**
   Although Maximizing FOV is more relevant to broadcasting; in the near future it will find its way to cinema as well in camera manufacturers and in RED V-Raptor firmware updates.

7 Conclusion

The black and white video frames are converted to colorized video frames. Also, the colorized videos are trained and tested by various video enhance AI models models Gaia High Quality 4K rendering, Theia fine Tune detail 4K not rendered and Theia Fine Tune Fidelity: 4K not rendered and video denoise AI models models Standard, clear, lowlight, severe noise and Raw. The upscaled and colorized video is also trained and tested using denoise video enhance AI and video denoise AI models. The results of each model are stored for comparison. From the stored results best video enhance AI model and the best video denoise AI models is selected. Lowlight AI model and Gaia high quality 4K rendering are selected and used in FVAF to produce high standard video for Forensic Analysis. We run this model using GPU to efficiently pre-process the video. By this framework, we increase the resolution, colorize, upscale, and denoise the video footages to successful investigations in the forensic field.

References

1. Li S, Li W, Cook C, Zhu C, Gao Y (2018) Independently recurrent neural network (IndRNN): building a longer and deeper RNN. arXiv:1803.04831 [cs.CV]
2. Shi W, Caballero J, Huszár F, Totz J, Aitken AP, Bishop R, Rueckert D, Wang Z Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. In: CVPR 2016

3. Kohonen T (1982) Self-organized formation of topologically correct feature maps. Biol Cybern 43(1):5969. https://doi.org/10.1007/bf00337288. S2CID 206775459
4. https://journalofbigdata.springeropen.com/articles/10.1186/s40537-021-00444-8
5. Bao W (2019) Depth-aware video frame interpolation
6. Chu M, Xie Y, Leal-Taixé L, Thuerey N Temporally coherent GANs for video super-resolution (TecoGAN)
7. Ronneberger O, Fischer P, Brox T U-Net: convolutional networks for biomedical image segmentation
8. Sun W, Sun J, Zhu Y, Zhang Y (2020) Video super-resolution via dense non-local spatial-temporal convolutional network (DENSENet). Neurocomputing
9. https://www.topazlabs.com/learn/introducing-video-enhance-ai-v2-3-frame-rate-conversion-and-slow-motion/
10. Karras T, Aila T, Laine S, Lehtinen J Progressive growing of GANs for improved quality, stability, and variation
11. https://github.com/LongguangWang/SOF-VSR
12. Zhang H, Goodfellow I, Metaxas D, Odena A Self-attention generative adversarial networks. Last revised 14 Jun 2019 (version, v2)
13. Bianchi FM, Maiorino E, Kampffmeyer MC, Rizzi A, Jenssen R (2017) An overview and comparative analysis of recurrent neural networks for short term load forecasting
14. Ulf E (2002) Radial basis function networks (RBFN) January 2002
15. Popescu MC, Balas VE, Perescu-Popescu L, Mastorakis N (2009) Multilayer perceptron and neural networks. WSEAS Trans Circ Syst 8(7):579–588
16. García-Lamont F, Cervantes J, López-Chau A et al (2020) Color image segmentation using saturated RGB colors and decoupling the intensity from the hue. Multimed Tools Appl 79:1555–1584. https://doi.org/10.1007/s11042-019-08278-6
17. Li C, Wang Y, Zhang X, Gao H, Yang Y, Wang J (2019) Deep belief network for spectral–spatial classification of hyperspectral remote sensor data. Sensors 19:204. https://doi.org/10.3390/s19010204
18. Broomhead DS, Lowe D (1988) Radial basis functions, multi-variable functional interpolation and adaptive networks. Technical report RSRE 4148. Archived from the original on April 9, 2013
19. Michelucci U (2022) An Introduction to Autoencoders. https://doi.org/10.48550/arXiv.2201.03898
20. Bao W, Lai W, Zhang X, Gao Z, Yang M (2021) MEMC-Net: motion estimation and motion compensation driven neural network for video interpolation and enhancement. IEEE Trans Pattern Anal Mach Intell 43(3):933–948. https://doi.org/10.1109/TPAMI.2019.2941941
21. Hinton G (2009) Deep belief networks. Scholarpedia 4(5):5947. Bibcode: 2009SchpJ...4.5947H. https://doi.org/10.4249/scholarpedia.5947
22. Rota C, Buzzelli M, Bianco S, Schettini R (2023) Video restoration based on deep learning: a comprehensive survey. Artif Intell Rev 56(6):5317–5364. https://doi.org/10.1007/s10462-022-10302-5
23. Sun W, Sun J, Zhu Y, Zhang Y (2020) Video super-resolution via dense non-local spatial-temporal convolutional network. Neurocomputing 403:1–12. https://doi.org/10.1016/j.neucom.2020.04.039
24. https://www.topazlabs.com/topaz-video-ai
25. Yang R, Timofte R, Zheng M, Xing Q NTIRE 2022 challenge on super-resolution and quality enhancement of compressed video: dataset, methods and results
26. Zhang H, Goodfellow I, Brain G, Odena A Self-attention generative adversarial networks
27. Creswell A, White T, Dumoulin V, Arulkumaran K, Sengupta B, Bharath AA Generative adversarial networks: an overview. IEEE
28. Radford A, Metz L, Chintala S Unsupervised representation learning with deep convolutional generative adversarial networks
29. Gulrajani I, Ahmed F, Arjovsky M, Dumoulin V, Courville A Improved training of Wasserstein GANs

30. https://www.topazlabs.com/VideoEnhanceAI/VideoQualitySoftware/TopazLabs
31. Vincent P, Larochelle H (2010) Stacked denoising autoencoders: learning useful representations in a deep network with a local denoising criterion. J Mach Learn Res. https://imagine-4d.com/multimmersion/
32. Rota C (2022) Video restoration based on deep learning: comprehensive survey. Springer
33. Schmidhuber J (1993) Habilitation thesis: system modeling and optimization. Page 150 ff demonstrates credit assignment across the equivalent of 1,200 layers in an unfolded RNN