Proceedings of the 4th International Conference on Mechatronics and Smart Systems
DOI: 10.54254/2755-2721/2026.33918

# *Video Course Super-Resolution Method Based on Scene Degradation Modeling*

**Guanting Wu**

*Shenzhen College of International Education, Shenzhen, China*
*waguating@126.com*

***Abstract.*** Video courses have become an important teaching medium due to the popularity of online education. However, affected by network bandwidth fluctuations and multiple transcoding compressions on the platform side, videos of online course often suffer from blurring, blocking artifacts and other quality degradation problems, which seriously affect the readability of key contents such as courseware text, formulas and blackboard writing. Most of the existing video super-resolution methods are designed for natural scenes, with single degradation assumptions and without special optimization for the content structure of instructional vides, and thus it is difficult to achieve ideal restoration effects. To solve this problem, this paper proposes a video super-resolution enhancement method oriented to online course scenarios. First, based on the structural characteristics of actual instructional videos, a special dataset is constructed, and a scene-based degradation synthesis strategy is designed, which highly restores the real network transmission and low-bitrate transcoding chain. Secondly, a readability-aware fine-tuning strategy is proposed by introducing structure-sensitive region masks and edge-preserving loss to guide the super-resolution network to focus on restoring of high-frequency information such as text strokes and formula boundaries. Experimental results reveal that, compared with existing mainstream super-resolution baseline models, the proposed method can achieve better performance in both subjective visual quality and objective evaluation metrics, and can effectively enhance the visual clarity and content readability of online course videos in complex compression scenarios.

***Keywords:*** Video Super-Resolution, Online Education, Degradation Modeling, Readability Awareness, Deep Learning

## 1. Introduction

With the popularization of online teaching, recorded courses, and live classrooms, course videos have become an important medium for the dissemination of teaching content and the accumulation of learning resources. Limited by network bandwidth fluctuations, platform transcoding, and multiple compressions, blur, blocking artifacts, ringing artifacts, and detail loss are commonly present in online course videos. These degradations directly affect the recognizability of courseware text, formula symbols, subtitle content, and local blackboard regions, thereby reducing the readability of course content and learning efficiency.

© 2026 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2755-2721/2026.33918

Super-resolution recovers high-resolution details from low-resolution inputs and is a classic research direction in image enhancement and video restoration. In recent years, deep learning has driven rapid development in super-resolution technology. From early convolutional reconstruction methods to approaches based on adversarial learning and realistic degradation modeling, models have achieved significant progress in visual quality and generalization ability [1-4]. When dealing with videos, using time-related information makes the results better. This allows video super-resolution to be used in many areas, such as security, improving movies, and online videos [5, 6].

But, most super-resolution techniques known today are made for natural images or general videos. The way these methods assume a video is damaged and the way they build it up do not really work well with online course videos. Online course videos have screen recordings, PPT pages, formula texts, and even some parts of the board. Plus, how these videos are coded and damaged is different because of live coding, changing bitrates, and codes from the platform. This makes the damage process more focused on compression and reacts to the content. In cases like this, only focusing on the whole picture is not enough. Because for teaching videos, the text structure, the formula edges, and the key information areas are more important.

Some super-resolution studies focus on a specific content area. These studies show that designing new data, damage, and training models to fit the situation works better than using the usual general super-resolution models. For example, research on animated videos. Creating new data and damage models for the situation has helped fix the video damage [7, 8]. Plus, research on online streaming points out that when the network changes, the video stream's bitrate and damage changes [9]. This makes it hard to use the usual fixed damage model. The above studies give us important ideas for repairing online course videos.

From the above observations, this paper studies the quality restoration task of online course videos, and proposes a super-resolution enhancement method for the online course scenario. Starting from the real degradation characteristics of online course videos, this paper constructs a course video dataset, designs a degradation generation strategy close to low-bitrate transmission and platform transcoding processes, and conducts targeted fine-tuning and optimization of existing super-resolution models to improve the restoration quality of courseware text, formula edges, and key teaching areas.

## 2. Related work

Single image super-resolution has always been a core direction of super-resolution research. SRCNN [1] introduced the convolutional neural network structure for the first time in a single image super-resolution task, verifying the effectiveness of the end-to-end deep learning framework. Subsequently, ESRGAN [2] greatly improved the perceptual quality of the reconstruction results through the way of residual generative adversarial learning, further promoting the transition of super-resolution research from pixel-level reconstruction to visual authenticity optimization.

With the deepening of continuous research, some scholars have begun to pay attention to complex degradation problems in real scenes. Real-ESRGAN [3] utilizes higher-order degradation processes to generate training samples closer to the real world, thereby enhancing the model's generalization ability under blur, noise and compression artifacts. BSRGAN [4] further proposes a more practical blind super-resolution degradation model, providing an effective degradation synthesis solution for real image restoration. Such works show that in real applications, degradation modeling itself is often as important as network structure.

Compared with single image super-resolution, video super-resolution can use the temporal correlation between adjacent frames to recover more detailed information. BasicVSR [5] constructs

DOI: 10.54254/2755-2721/2026.33918

a structurally simple but effective video super-resolution baseline through bidirectional propagation and feature alignment. On this basis, RealBasicVSR [6] further targets real video degradation scenarios, and discusses the trade-off relationship between restoration quality, efficiency and complexity, providing an important benchmark for real video enhancement tasks.

Although the above methods have obtained good results in general video restoration, their training settings still mainly revolve around natural videos. When the target scene has obvious content structural features and compressed transmission links, it may still be difficult to obtain optimal results by relying solely on general video super-resolution models.

In the past few years, researchers have begun to pay attention to the restoration problems of super-resolution in specific content fields, among which animation and cartoon videos are representatives. Take animation video as an example, AnimeSR [7] constructed the AVC dataset for real animation video super-resolution tasks, and improved the restoration effect of animation videos through degradation learning and efficient network design. In addition, VQD-SR [8] further adopts a data-driven vector quantization degradation modeling method to more accurately characterize the complex degradation distribution of animation videos.

For online streaming scenarios, existing work has pointed out that the degradation of real-time video streams is not static and fixed but dynamically fluctuates with changes in network bandwidth and streaming encoding parameters. Yin et al. [9] constructed the LDV-WebRTC dataset and confirmed that videos under real-time transmission environments have obvious dynamic variation characteristics in bitrate, quantization parameters, and visual quality. This suggests that traditional super-resolution methods based on fixed degradation assumptions often suffer from poor adaptation in streaming scenarios.

In comparison, there is relatively little research on super-resolution for online course videos. Online course videos have higher requirements for the readability of text, formulas, interface layout, and key teaching content than ordinary streaming videos. The restoration goals of such videos are not only to improve subjective visual clarity but also to ensure the accurate preservation of teaching information. Although the existing super-resolution research on specific scenarios provides some references for the restoration of compression degradation, there is still space for further research in course scenario adaptation and teaching content-friendly modeling, which is also the starting point of this work.

## 3. Methodology

This article suggests a framework for improving online course video quality using super-resolution enhancement methods. Figure 1 shows the full process. The framework uses high-quality course videos to create training data and employs a degradation synthesis strategy designed for online course settings to produce low-quality input samples. It also fine-tunes scenario-specific versions of existing super-resolution backbone networks to finally output enhanced video frames. At the same time, to make the system more practical in real teaching situations, this article also adds modules for real-time transcription and important content summarizing after class to help with course searching and reviewing.

DOI: 10.54254/2755-2721/2026.33918

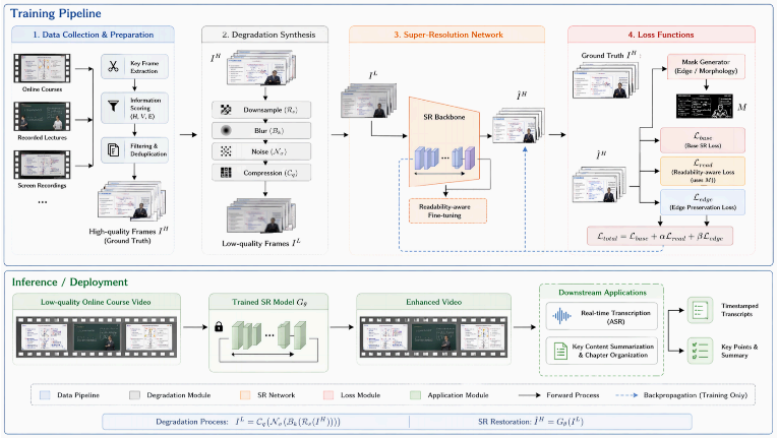


Figure 1. The overall process for super-resolution enhancement of video quality restoration

Let the high-quality course video frame be $I^H$ (I superscript H), the degraded low-quality input be $I^L$ (I superscript L), and the super-resolution restoration network be $G\theta(\cdot)$, then the restoration result can be expressed as equation (1):

$$\hat{I}^H = G\theta(I^L) \tag{1}$$

where $\hat{I}^H$ denotes the reconstructed high-resolution image output by the model. The core of the proposed method lies in constructing data samples more consistent with online course scenarios, designing degradation chains closer to the real transmission process, and enhancing the restoration effect of text, formulas, and blackboard writing areas through a readability-aware perceptual training strategy.

### 3.1. Course video dataset construction

To better adapt the model to the content distribution of online course videos, this paper first constructs a course video dataset oriented towards online course scenarios. Unlike natural videos, online course videos usually have obvious layout structures and content region differences, such as titles, main text, formula boxes, code blocks in PPT pages, as well as text strokes and teacher writing areas in blackboard videos. Directly sampling the original video frame-by-frame easily introduces a large number of duplicate frames, low-information frames, and over-compressed frames, thus reducing training efficiency. Therefore, this paper adopts the process of "video collection – keyframe extraction – information filtering – paired construction" to generate training samples. First, original videos are collected from open courses, recorded teaching videos, and high-quality screen-recorded courses, and uniformly transcoded into higher-quality formats as candidate sources. Subsequently, the original frame set $\{I_i\}^k_{i=1}$ is extracted according to a fixed interval or keyframe strategy. To reduce the proportion of redundant frames, this paper calculates an information scoring function for each frame as equation (2):

$$S(I_i) = \alpha H(I_i) + \beta V(I_i) + \gamma E(I_i) \tag{2}$$

where $H(I_i)$ represents the image information entropy, $V(I_i)$ represents the variance of the Laplacian response used to characterize sharpness, and $E(I_i)$ represents the edge pixel density used to describe the richness of structural content such as text, formulas, and blackboard writing; $\alpha$, $\beta$, $\gamma$

DOI: 10.54254/2755-2721/2026.33918

are weighting coefficients. Based on the scoring results, keyframes with higher information content and significant visual differences in each video segment are retained, while pure blank pages, duplicate pages, and pages with extremely low information are removed to improve the effectiveness of training samples.

When constructing training pairs, the filtered high-quality frames are used as approximate real high-resolution samples and corresponding low-quality samples are generated through subsequent degradation models, thereby forming paired data ($I^H$, $I^L$). For video clip-level training, the same degradation parameters can also be maintained on adjacent frames to simulate the real transmission process with relatively stable bitrate and image quality over a short period.

### 3.2. Degradation synthesis model for online course scenarios

The quality degradation of real online course videos is usually not caused by a single factor, but is the result of the combined effects of resolution reduction, blur disturbance, and encoding compression. Especially for PPT text, formula edges, and subtitle areas, low bitrate transmission and platform transcoding often lead to obvious detail loss, edge blurring, and compression artifacts. To this end, this paper constructs a degradation synthesis model oriented to online course scenarios, used to simulate the typical distortion forms of online course videos during the real transmission process.

Given a high-quality input $I^H$, its corresponding low-quality sample $I^L$ is defined as equation (3):

$$I^L = C_v(N\sigma(B^k(R(I^H)))) \tag{3}$$

where R represents the scaling operation, s is the scaling ratio; $B^k$ represents the blur operation, k is the blur intensity; $N\sigma$ represents noise disturbance, $\sigma$ is the noise intensity; $C_v$ represents the compression process, q is the compression quality parameter.

This degradation model mainly includes the following three parts.

(1) Resolution reduction. Online course videos usually undergo resolution changes during acquisition, uploading, and playback, so this paper first randomly scales the high-quality image to simulate detail loss under different playback resolutions. This operation is mainly used to generate basic low-resolution inputs and enhance the model's adaptability to course videos of different sizes.

(2) Blur and noise disturbance. To simulate slight quality damage generated during screen recording, video sampling, and transmission, blur and noise disturbance are superimposed on the scaled image. Among them, the blur operation is mainly used to depict the softening phenomenon of text strokes and formula boundaries, and the noise term is used to enhance the model's robustness to slight acquisition noise and encoding disturbance.

(3) Compression degradation. Compression artifacts are one of the main sources of distortion in online course videos. To this end, a lossy compression process is added at the end of the degradation chain to simulate the blocking artifacts, ringing artifacts, and high-frequency detail loss caused by low bitrate transmission or platform transcoding. Through this design, the generated low-quality samples can more closely approximate common degradation forms in real online course videos.

### 3.3. Readability-aware fine-tuning strategy

Considering that key information in online course videos is mainly concentrated in text, formulas, and blackboard writing areas, this paper introduces a readability-aware fine-tuning strategy to the basic super-resolution model to enhance the model's ability to restore teaching content. Specifically,

DOI: 10.54254/2755-2721/2026.33918

a structure-sensitive region mask M is first generated based on edge detection and morphological processing, where M(x) = 1 indicates that pixel x belongs to text or key structural areas, and M(x) = 0 indicates ordinary areas.

On this basis, the weighted reconstruction loss is defined as equation (4):

$$L_r^{ead} = (1/N) \times \Sigma_x (M(x) + \lambda(1 - M(x))) \times |\hat{I}^H(x) - I^H(x)| \quad (4)$$

where $\hat{I}^H$ is the model output, $I^H$ is the corresponding high-resolution real image, N represents the total number of pixels, and $\lambda \in (0, 1)$ is the weight coefficient for ordinary areas. Through this design, the model pays more attention to the reconstruction error of key areas such as text, formulas, and blackboard writing during the training process, thereby improving the readability restoration effect of course content.

To further enhance the clarity of edge structures, an edge-preserving loss is introduced as equation (5):

$$L^{edge} = ||\nabla\hat{I}^H - \nabla I^H||_1 \quad (5)$$

where $\nabla(\cdot)$ represents the image gradient extraction operator. This loss is used to constrain the model to restore high-frequency structural information such as text strokes, formula boundaries, and blackboard lines.

Finally, the training objective of this paper is defined as equation (6):

$$L^{totaL} = L^{ba*e} + \alpha L_r^{ead} + \beta L^{edge} \quad (6)$$

where $L^{ba*e}$ represents the original training objective of the basic super-resolution model, and α and β are weight coefficients. Through this lightweight design, targeted adaptation for online course scenarios is achieved without making substantial modifications to the original network structure.

## 4. Experiments

### 4.1. Experimental settings and results

Public recorded courses, screen-recorded courses, and blackboard teaching videos were collected to construct a course video dataset. The data was divided into training, validation, and testing sets after keyframe filtering. The test set includes two parts: synthesized degraded data and real online course videos, where synthesized data is used for full-reference evaluation and real data is used for no-reference evaluation. Comparative methods include Bicubic, Real-ESRGAN, and RealBasicVSR.

Table 1 shows the results of each method on the test set. It can be seen that the proposed method achieved good results on both the synthesized test set and the real test set, indicating that the degradation modeling and training strategy proposed in this paper for online course video enhancement are superior to general baselines.

Table 1. Experimental results

| Method | PSNR↑ | SSIM↑ | LPIPS↓ |
|---|---|---|---|
| Bicubic | 27.82 | 0.811 | 0.184 |
| Real-ESRGAN | 28.41 | 0.829 | 0.149 |
| RealBasicVSR | 28.73 | 0.838 | 0.136 |
| Experiment | 29.76 | 0.865 | 0.108 |

DOI: 10.54254/2755-2721/2026.33918

To further illustrate the enhancement effect of the super-resolution network on online course materials, the following cases are selected, covering multiple subjects such as biology, geography, and history. The content includes common course scenes such as legends, text, and blackboard writing.

## 4.2. Ablation experiments

To test how well each module works as described in this paper, we did ablation tests using the same training setup. Starting with the basic super-resolution model, we added degradation modeling that works with online courses and readability perceptual loss. We tested our results using PSNR, SSIM, and LPIPS on the test set.

Table 2 shows these results. After bringing in scenario-based degradation modeling, the model performed better, showing that this method can better copy the real-life degradation of online course videos. When we brought in readability perceptual loss, the model was even better at restoring text, formulas, and blackboard writing, which are key areas.

Table 2. Ablation experiment results

| Method | PSNR↑ | SSIM↑ | LPIPS↓ |
|---|---|---|---|
| Basic model | 28.62 | 0.817 | 0.134 |
| + Degradation modeling | 29.11 | 0.851 | 0.114 |
| + Readability perceptual optimization | 29.76 | 0.865 | 0.108 |

## 5. Conclusion

This paper addresses issues such as blurring, compression artifacts, and detail loss that easily occur in online course videos under low bitrate transmission and platform transcoding conditions, and proposes a video course super-resolution enhancement method based on scenario degradation modeling. By constructing a dataset oriented to online course scenarios and designing a degradation synthesis strategy close to the real transmission process, and fine-tuning the basic super-resolution model combined with readability-aware perceptual loss, the restoration effect of courseware text, formula edges, and blackboard writing areas in course videos is improved.

From the experimental results, our method performs better than the comparative methods on both objective evaluation metrics and subjective visual effects. It can effectively improve the clarity and content readability of online course videos in complex compression scenarios. Future work will further increase the scale of course video data and improve the quality evaluation methods in real scenes to enhance the application value of the method in real online teaching scenes.

## References

[1] Dong C, Loy CC, He K, Tang X. Image super-resolution using deep convolutional networks. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2015; 38(2): 295-307.

[2] Wang X, Yu K, Wu S, Gu J, Liu Y, Dong C, et al. ESRGAN: Enhanced super-resolution generative adversarial networks. In: Proceedings of the European Conference on Computer Vision Workshops (ECCV Workshops); 2018.

[3] Wang X, Xie L, Dong C, Shan Y. Training real-world blind super-resolution with pure synthetic data. arXiv preprint arXiv: 2107.10833; 2021.

[4] Zhang K, Liang J, Van Gool L, Timofte R. Designing a practical degradation model for deep blind image super-resolution. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV); 2021.

DOI: 10.54254/2755-2721/2026.33918

[5] Chan KCK, Wang X, Yu K, Dong C, Loy CC. BasicVSR: The search for essential components in video super-resolution and beyond. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR); 2021. p. 4947-4956.
[6] Chan KCK, Zhou S, Xu X, Loy CC. Investigating tradeoffs in real-world video super-resolution. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR); 2022. p. 5962-5971.
[7] Wu Y, Wang X, Li G, Shan Y. AnimeSR: Learning real-world super-resolution models for animation videos. In: Advances in Neural Information Processing Systems (NeurIPS); 2022.
[8] Tuo Z, Yang H, Fu J, Dun Y, Qian X. Learning data-driven vector-quantized degradation model for animation video super-resolution. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV); 2023.
[9] Yin G, Qu Z, Jiang X, Jiang S, Han Z, Zheng N, et al. Online streaming video super-resolution with convolutional look-up table. IEEE Transactions on Circuits and Systems for Video Technology. 2024.