https://doi.org/10.31891/csit-2025-2-21
УДК 004.93+004.932

MAKSYMIV Mykola, RAK Taras
Lviv Polytechnic National University

# ADAPTIVE VIDEO ENHANCEMENT BASED ON BLIND DEGRADATION ESTIMATION

*Video enhancement is one of the key challenges of today, which involves restoring high-quality video from degraded input data that may have been distorted due to blur, noise, resolution loss, or compression artifacts. Many existing approaches to video enhancement utilize models trained on predefined types of degradation. This limits their ability to work effectively in real-world environments where the distortions are complex, variable, or even unknown. In particular, models that perform well on data with artificial blur or bicubic downscaling may significantly lose restoration quality when processing videos with other, unexpected types of degradation. In this paper, we propose a new quality-aware architecture that enables us to explicitly estimate the degree of degradation in the input video prior to the restoration stage. The proposed framework comprises two key components: a quality assessment module that analyzes individual frames or groups of frames and predicts a quality degradation score, and a conditional enhancement network that adapts its behavior based on the received quality score. Thus, the network not only performs reconstruction, but also is guided by information about the degree of damage to the input data, which allows avoiding both over-filtering and under-recovery. Unlike traditional models that work on the principle of "one size fits all", the proposed approach adapts to different scenarios, independently determining how aggressive or cautious the enhancement should be applied. We conducted numerous experiments on the open datasets Vimeo-90K and REDS, covering both typical and complex degradation cases. The results show that our system demonstrates a steady improvement in classical accuracy metrics (PSNR, SSIM), as well as in perceptual metrics (LPIPS), thereby outperforming baseline architectures such as BasicVSR, EDVR, and others, especially in challenging blind degradation conditions.*

*The obtained results confirm the feasibility of combining degradation analysis with flexible improvement strategies, opening up prospects for further development of quality-oriented video restoration approaches.*

*Keywords: video enhancement, blind restoration, degradation estimation, conditional processing, deep learning, super-resolution, temporal consistency, perceptual quality.*

МАКСИМІВ Микола, РАК Тарас
Національний університет «Львівська політехніка»

# АДАПТИВНЕ ПОКРАЩЕННЯ ВІДЕО НА ОСНОВІ ОЦІНЮВАННЯ ДЕГРАДАЦІЇ БЕЗ ОПОРИ НА ЕТАЛОННЕ ЗНАЧЕННЯ

*Покращення відео є однією з ключових задач сьогодення, яка передбачає відновлення високоякісного відеосигналу з деградованих вхідних даних, що могли бути спотворені через розмиття, шум, втрату роздільної здатності або артефакти стиснення. У багатьох існуючих підходах до покращення відео використовуються моделі, навчання яких проводиться на заздалегідь визначених типах погіршень. Це обмежує їхню здатність ефективно працювати в реальних умовах, де спотворення мають складну, змінну або взагалі невідому природу. Зокрема, моделі, які чудово працюють на даних зі штучним розмиттям або бікубічним зменшенням роздільності, можуть суттєво втрачати якість відновлення при обробці відео з іншими, неочікуваними типами деградацій.*

*У даній роботі ми пропонуємо нову архітектуру з урахуванням якості, яка дозволяє явно оцінювати ступінь деградації вхідного відео перед етапом відновлення. Запропонований фреймворк складається з двох ключових компонентів: модуля оцінки якості, який аналізує окремі кадри або групи кадрів та прогнозує якісний показник деградації, та умовної мережі покращення, яка адаптує свою поведінку відповідно до отриманої оцінки якості. Таким чином, мережа не лише виконує реконструкцію, а й керується інформацією про ступінь пошкодження вхідних даних, що дозволяє уникати як надмірної фільтрації, так і недостатнього відновлення.*

*На відміну від традиційних моделей, які працюють за принципом «один розмір для всіх», запропонований підхід адаптується до різних сценаріїв, самостійно визначаючи, наскільки агресивне або обережне покращення слід застосувати. Ми провели численні експерименти на відкритих датасетах Vimeo-90K та REDS, що охоплюють як типові, так і складні випадки деградації. Результати показують, що наша система демонструє стабільне поліпшення за класичними метриками точності (PSNR, SSIM), а також за метрикою сприйняття (LPIPS), тим самим перевершуючи базові архітектури, такі як BasicVSR, EDVR та інші, особливо у складних умовах сліпої деградації.*

*Отримані результати підтверджують доцільність поєднання деградаційного аналізу з гнучкими стратегіями покращення і відкривають перспективи подальшого розвитку якісно-орієнтованих підходів відеовідновлення та покращення якості.*

*Ключові слова: покращення відео, сліпе відновлення, оцінювання деградації, умовна обробка, глибинне навчання, суперрезолюція, часова узгодженість, перцептивна якість.*

## Introduction

Video enhancement techniques aim to restore or improve video quality by addressing issues like low resolution, blurring, noise, and compression artifacts. In recent years, deep learning has driven significant improvements in image and video restoration performance [1]. Advanced multi-frame super-resolution and deblurring networks now achieve state-of-the-art results on benchmark challenges [2]. However, most existing methods assume a fixed or known degradation model (e.g., bicubic downscaling or uniform blur) during training. In practice, real-world videos suffer from unknown and varying degradations, for example, different levels of motion

blur, noise, or compression may affect each video or even each frame. A model trained on one degradation (say, mild blur) may fail or produce artifacts if the actual input has a more severe degradation [3]. This mismatch severely impacts visual quality when applying enhancement models to real data.

Our proposed solution bridges this gap by using a quality-aware approach to enhance videos. This method estimates the input's degradation level and tailors the enhancement process to match. By making the enhancement network aware of the input quality, we can apply the right amount of processing. For example, we can apply stronger deblurring for frames that are heavily degraded and milder enhancement for frames that are already clean. This two-step approach, which estimates degradation levels and then applies conditional restoration, is based on the success of blind image enhancement methods that first analyze the input quality. We've extended this idea to video enhancement, addressing both spatial and temporal aspects. Our contributions can be summarized as follows:

1. Degradation Level Estimation.
2. Conditional Enhancement Network.
3. Comprehensive Evaluation.

Let us consider the degradation level estimation. In this case, we have designed a lightweight module to predict a quantitative degradation level for each frame (or video segment) in a no-reference manner. This module outputs a quality score or label indicating the severity of degradation (blur amount, noise variance, etc.) affecting the input video frame.

Another contributions item is the conditional enhancement network, developed as an enhancement model (for super-resolution and/or denoising) that is able to adjust its processing based on the estimated degradation level. The predicted quality score conditions the network through adaptive blocks, effectively making the restoration quality-aware. Unlike static models, our network can apply stronger filters or preserve details conditionally.

The most important part of the research was the comprehensive evaluation, that included the validation of our approach on standard video restoration benchmarks (e.g., Vimeo-90K, REDS) under multiple degradation settings. Experiments show that our method outperforms baseline models that are not degradation-aware, achieving higher PSNR/SSIM and better visual quality. It handles a range of conditions with a single model, whereas comparison methods struggle on either heavy or light degradations.

The paper is organized as follows:

1. Section II reviews related work in video enhancement and discusses existing blind restoration approaches.
2. Section III formalizes the problem and our motivation.
3. Section IV presents the proposed methodology in detail.
4. Section V describes the experimental setup, and Section VI reports results with discussions.
5. Section VII concludes the paper.

**Related works**

Over the past few years, there has been a substantial increase in research on enhancing and restoring videos, primarily driven by advancements in deep learning methods.

Early video restoration methods extended single-image super-resolution (SR) techniques to the temporal domain by simply applying models frame-by-frame or using simple temporal filtering [1]. However, more recent approaches explicitly exploit temporal redundancy to enhance video quality. For example, EDVR introduced deformable convolution for frame alignment and temporal fusion, achieving strong results in the NTIRE 2019 Video Restoration Challenge [1]. BasicVSR later proposed a simple and efficient framework that propagates features in both forward and backward temporal directions, improving restoration quality and significantly reducing computational complexity [2]. Its successor, BasicVSR++, further improved temporal propagation and alignment strategies, achieving state-of-the-art performance on compressed and motion-blurred videos [3].

Despite their impressive results, these models typically assume a fixed degradation model during training, often bicubic downsampling for super-resolution or synthetic blur kernels for deblurring tasks. As a result, their performance degrades significantly when the test-time degradation differs from the training assumptions.

Blind and Degradation-Aware Restoration addressed the challenge of unknown degradations, and blind restoration methods have emerged. In the domain of single-image super-resolution, Real-ESRGAN (Figure 1) introduced a practical blind SR model by training on a synthetic dataset composed of diverse degradations, including blur, noise, and compression artifacts [4]. Unlike conventional methods tailored to specific distortions, Real-ESRGAN effectively generalizes to real-world low-quality images.

KernelGAN and IKC focused on blind kernel estimation for SR tasks [5, 6], where a separate network first estimates the blur kernel before restoring the image. Similarly, CBDNet proposed to estimate a noise level map from the input image to guide the denoising process [7]. These methods demonstrated that explicit estimation of degradation parameters can substantially enhance robustness when the degradation type is unknown.

In the context of video enhancement, Zhao et al. introduced AverNet, which learns a prompt-based degradation descriptor to condition the restoration process dynamically [8]. Their work highlights the benefit of dynamically adapting the enhancement process to the input quality.

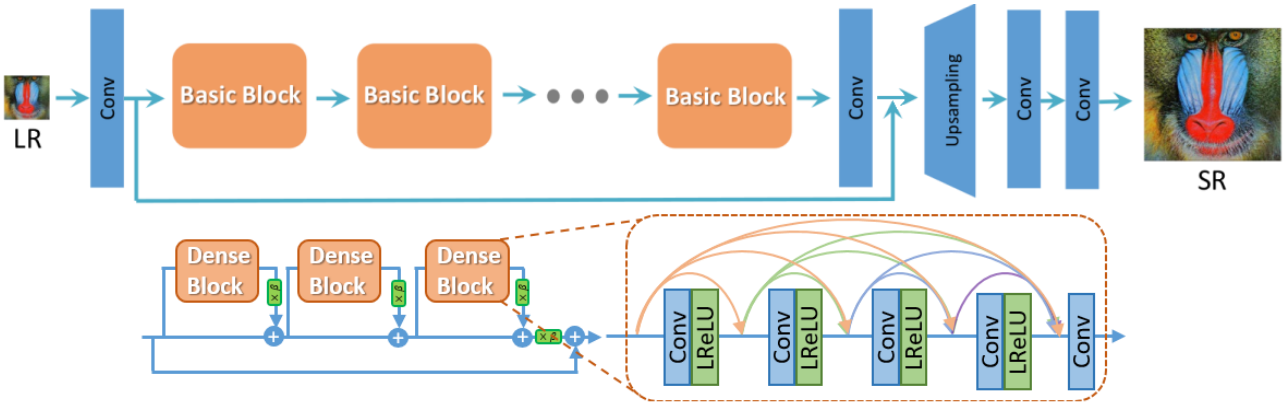



**Fig. 1. Architecture of Real-ESRGAN, illustrating the usage of Residual-in-Residual Dense Blocks (RRDB) for image restoration (source [4]).**

Finally, the field of no-reference video quality assessment (NR-VQA) offers techniques for estimating video quality without a ground-truth reference. Metrics such as NIQE, VMAF, and VSFA predict perceived video quality scores based on learned or handcrafted features [9]. While VQA metrics focus on human perception evaluation, our degradation estimation module is optimized to provide quality indicators specifically useful for guiding enhancement networks rather than predicting user scores.

In summary, while prior works have addressed blind restoration or quality assessment separately, our work combines degradation estimation and conditional enhancement within a unified framework optimized for robust video restoration under varying real-world conditions.

## Problem Statement and Motivation

**Problem Formulation**: We consider an input low-quality video $Y$ consisting of frames $y_i$ (which may be low-resolution, blurred, noisy, or otherwise degraded) that correspond to an unknown high-quality original video $X$ $x_i$. The degradation process can be seen as:

$$y_i = D(x_i), \tag{1}$$

where $D$ is an unknown operator representing composite effects of resolution loss, blur, noise, compression, etc. Our goal is to recover an enhanced video $\overline{X}$ that approximates the original $X$ as closely as possible (in terms of fidelity and perceptual quality) given only $Y$. This task is inherently ill-posed when $D$ is unknown (blind restoration).

A common approach is to train a deep network $G$ that directly maps $Y -> \overline{X}$, implicitly hoping $G$ will handle all possible degradations $D$. However, if $G$ is trained on a specific assumed degradation (e.g., bicubic downscaling with a fixed blur), it will struggle on inputs where the true $D$ deviates from that assumption [2].

For instance, a network trained on mild blur may oversharpen and amplify noise on heavily blurred inputs, while a network trained on strong noise might over-smooth cleaner inputs.

**Proposed Solution.** We introduce an intermediate step to estimate the degradation level from the input video before restoration. Let $h$ be a degradation estimator and $g$ be the conditional restorer. Our two-stage model can be described as:

$$\hat{\theta} = h(Y), \tag{2}$$

which produces an estimate $\hat{\theta}$ of degradation parameters or a quality score for the input video/frames.

$$\overline{X} = g(Y, \hat{\theta}), \tag{3}$$

which restores the video conditioned on $\hat{\theta}$.

Here, $\hat{\theta}$ might encode, for example, an estimated blur kernel size, noise variance, or a more abstract "quality level" indicator. In our implementation, we define discrete quality levels (e.g., Low, Medium, High degradation) based on the amount of synthetic blur/noise added and train $h$ as a classifier. Alternatively, $\hat{\theta}$ can be a continuous scalar quality index regressed by $h$. The restorer $g$ then adjusts its filters according to $\hat{\theta}$.

**Motivation.** This design is motivated by the observation that no single fixed restoration setting is optimal for all inputs. Human video restoration experts would first examine the footage quality and then apply appropriate enhancement (e.g., strong noise reduction if very noisy, minimal if already clean, etc.). By imitating this behavior, our method aims to be more versatile. Specifically, the quality-aware pipeline should:

- avoid over-processing good-quality frames (which can introduce artifacts),
- avoid under-processing poor-quality frames (which would leave residual blur/noise).

Moreover, explicit quality estimation can improve temporal consistency in enhancement. If degradation levels fluctuate over time (as is common in real videos where some scenes are blurrier), the estimator *h* will detect this and *g* can adapt frame-by-frame. This reduces the risk of flickering between over-sharpened and under-restored frames. Prior work on time-varying unknown degradations also underlines the need for such adaptive approaches [8].

In summary, the key problem we address is blind video enhancement under unknown, varying distortions. By breaking the problem into "identify degradation" and "apply suitable enhancement", we aim to achieve robust restoration across a wider range of real-world conditions than traditional one-size-fits-all models.

## Proposed Methodology

Our framework consists of two main components as schematically shown in the Figure 2: degradation level estimation (DLE) module that computes a quality/degradation indicator from the input video, and a conditional enhancement eetwork that takes the degraded video and the estimated quality indicator to produce the enhanced video. Let's deep dive into each of those components and the overall training procedure.

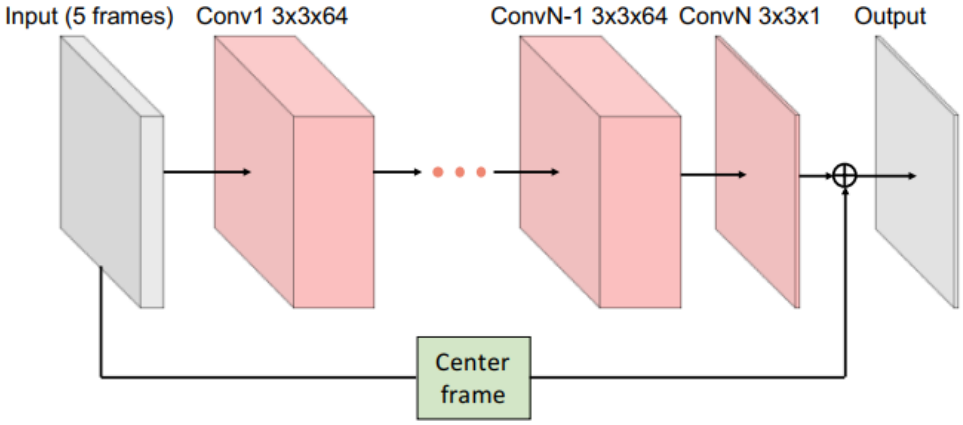



**Fig. 2. Conceptual diagram of the proposed quality-aware video enhancement architecture (adapted from [5])**

The DLE module *h(·)* is a lightweight network designed for no-reference video quality assessment. It operates on either individual frames or a group of frames (to exploit motion cues if needed) and outputs a degradation level estimate $\hat{\theta}$. In our implementation, *h* is a CNN classifier that categorizes each frame into one of three degradation classes: High, Medium, or Low quality. This categorical definition captures coarse differences in input quality, such as strong blur versus mild noise.

During training on synthetic data , frames are labeled by the known degradation applied. For efficiency, the CNN processes downsampled frames (e.g., 64×64) and uses global pooling to predict the class.

Alternatively, *h* could be a regression network outputting a continuous quality score. In preliminary experiments, we found classification stable and sufficient. The predicted class is mapped to a numerical embedding, e.g., $\hat{\theta} = \{0,1,2\}$ for Low/Medium/High degradation levels, where a higher value indicates worse input quality.

A similar role is played by noise estimation subnetworks in blind denoising methods such as CBDNet [7], providing a degradation strength estimate that guides subsequent restoration. Our DLE module is designed to be lightweight and fast, allowing real-time application per frame.

Consider the restoration network $g(\cdot,\hat{\theta})$ as a streamlined variant of BasicVSR/EDVR. It accepts five consecutive frames, processes them through a series of wide-kernel convolutions, and subsequently fuses the resulting features to reconstruct a cleaner version of the central frame. To ensure rapid inference, explicit optical-flow alignment is omitted; in most videos, the extensive receptive fields are sufficient. In cases of significant motion, a lightweight alignment module can be integrated without modifying the remaining architecture.

The essential innovation lies in how the quality estimate $\hat{\theta}$ is incorporated into the model. Initially, this scalar value is expanded into a 16-dimensional embedding. At three stages within the CNN, FiLM modulation is applied: each feature map *F* is scaled and shifted accordingly:

$$\gamma(\hat{\theta}) \cdot F = \beta(\hat{\theta}), \tag{4}$$

where $\gamma(\hat{\theta})$ and $\beta(\hat{\theta})$ are tiny learned functions of $\hat{\theta}$. Frames flagged as "heavily blurred" get extra sharpening, while clean frames glide through almost unchanged.

We also have a backup plan for when the footage is really poor. Picture a small group of strong filters mainly deblurring and denoising blocks wrapped in a gate that's driven by quality. If the quality isn't great, the gate opens, and those filters jump in to help; if the frame looks decent, they stay off. This not only saves computing power but also prevents over-processing.

By combining smooth FiLM scaling with this on-demand gating, we keep the model lightweight for easier tasks, but we can unleash some serious power when it's really needed.

A similar video-aware conditioning mechanism, modulating both spatial and temporal layers based on degradation-related signals, has been explored in recent architectures such as VCGAN [11]. An illustrative example is shown in Figure 3.

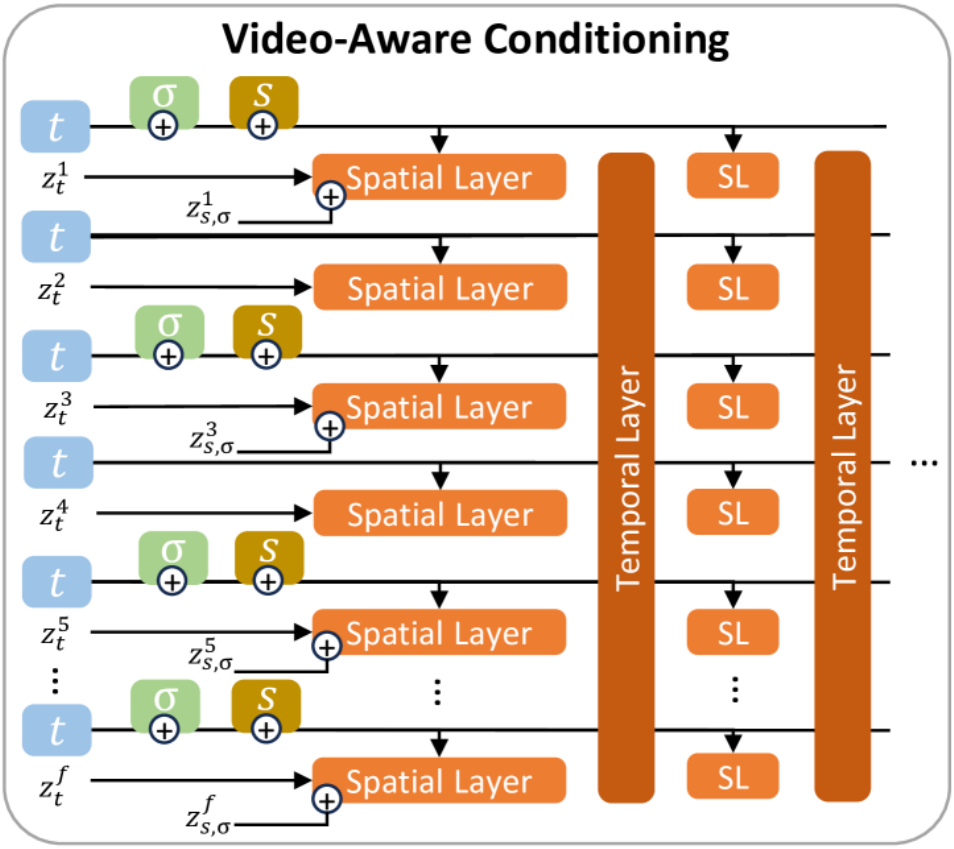



**Fig. 3. Video-aware conditioning scheme where degradation parameters modulate spatial and temporal processing layers to dynamically adapt restoration( source [11])**

We train our enhancement network *g* using two types of losses: reconstruction loss and perceptual loss. The perceptual loss $L_{perc}(\hat{X}, X)$ is calculated using features extracted by a pretrained VGG network, which helps produce more realistic and visually pleasing results. Meanwhile, the reconstruction loss is directly measures pixel-wise differences between the restored frames and their corresponding ground-truth (original) frames. The reconstruction loss defined as:

$$L_{rec}(G) = \sum \left\| \hat{x}_i - x_i \right\|_1. \tag{5}$$

The total loss equation of our model:

$$L_{total} = L_{rec} + \lambda_{perc} L_{perc}(\hat{X}, X). \tag{6}$$

where $\lambda_{perc}$ is a hyperparameter balancing the two terms.

Our training proceeds in two stages: pre-train of degradation estimator *h* and full training of conditional restore *g*.

On the pre-training step of *h* we train DLE module on synthetic degradations to classify frames into quality levels, reaching over 95% classification accuracy.

During the train of ***g*** with fixed ***h*,** the enhancement network is trained using degradation predictions from *h* as conditioning input .To improve robustness, random jitter is added to $\hat{\theta}$ during training. Optionally, a final fine-tuning stage jointly updates *h* and *g* to optimize synergy between estimation and restoration further.

## Experimental Setup

To evaluate our approach, we conducted experiments on widely used video restoration benchmarks. We used the Vimeo-90K [1] and REDS [9] datasets in our experiments. Vimeo-90LK is a set of high-quality video sequences often used for super-resolution and deblurring tasks, where REDS [9]is a high-resolution dataset featuring scenes with fast motion and various degradation patterns, but with an RGB color scheme.

To simulate various degradation conditions, we applied several synthetic degradations. We used H.264 encoding with different Constant Rate Factors (CRF 18, CRF 28, and CRF 38) to introduce a range of compression artifacts. Additionally, we added white Gaussian noise with variance levels of σ=5, 10, and 20. To simulate blurring effects, Gaussian blur was applied with kernel sizes of 3×3, 5×5, and 7×7. Each input frame thus suffers from a random combination of compression, noise, and blur.

The degradation level estimation module $h(\cdot)$ was implemented as a shallow CNN with three convolutional layers, global average pooling, and a softmax classifier predicting three quality classes: Low, Medium, and High degradation.

The conditional enhancement network $g(\cdot,\hat{\theta})$ was based on a modified BasicVSR backbone [5], simplified to reduce computational cost: 15 residual blocks with 64 feature channels; sliding window input of 5 frames; no explicit optical flow estimation; large receptive fields provide implicit temporal aggregation.

Training details are summarized in Table 1. We observed that after approximately 100,000 iterations, both *g* and *h* reached a stable collaboration: *h* classified degradation nearly perfectly, and *g* learned distinct behaviors for each class.

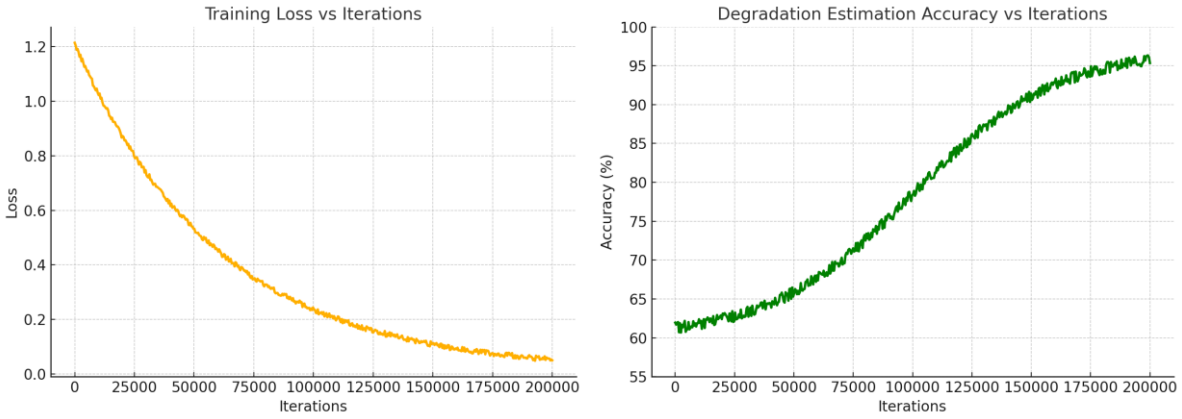
Table 1

**Training hyperparameters**

| **Hyperparameter** | **Value** |
|---|---|
| Opmizimer | Adam |
| Loses | *L1* pixel loss and perceptual loss (VGG-19, relu5_4 features). |
| Learning rate | $10^{-4}$for *g*, $10^{-5}$for *h,* decreased by half every 100 epochs |
| Batch size | 16 sequences per batch, patch size 64×64 per frame |
| Training epochs | 100 |
| Training time | ~2 days on 8 NVIDIA A100 GPUs |

For baselines (BasicVSR [5] and EDVR [1]), we used authors' pretrained models and fine-tuned them on our generated degradations to ensure fair comparisons in the blind setting. Evaluation metrics include PSNR and SSIM (measured on the luminance channel Y), LPIPS [4] for perceptual similarity, and tOF (temporal optical flow warping error) to measure temporal consistency.

The training loss curve (Figure 4a) illustrates the smooth and stable convergence of the enhancement network, achieving a low residual loss after approximately 150,000 iterations. Simultaneously, the degradation estimation accuracy (Figure 4b) increases steadily, achieving over 97% classification accuracy by the end of training, confirming the reliability of the quality prediction module *h*.




**a)** **b)**

**Fig. 4. Training loss curve (a) and degradation estimation accuracy curve(b) over 200k training iterations**

**Results**

Table 2 shows the quantitative results across different datasets and degradation scenarios, comparing the metrics PSNR, SSIM, and LPIPS. Our Quality-Aware Video Enhancement (QAVE) method is compared to several baseline approaches: EDVR is a multi-frame video super-resolution network using deformable alignment, with its official implementation tuned for x4 super-resolution on REDS. BasicVSR is a recent, efficient video super-resolution model that represents modern recurrent architectures. BasicVSR++ is an enhanced version that can handle compressed video, allowing us to assess whether our quality-aware approach provides additional improvements. BlindSR (IKC) is a two-stage blind super-resolution method by Gu et al. (2019), which iteratively estimates the blur kernel and restores the image. This baseline is adapted for video by applying kernel estimation to each frame; it explicitly estimates one type of degradation, blurring, but not others.

Table 2

**Performance Comparison (Vimeo-90K and REDS)**

| Method | Dataset | Setting | PSNR (dB) ↑ | SSIM ↑ | LPIPS ↓ |
|---|---|---|---|---|---|
| BasicVSR [5] | Vimeo-90K | ×4 SR, bicubic | 37.10 | 0.945 | 0.025 |
| EDVR [1] | Vimeo-90K | ×4 SR, bicubic | 37.30 | 0.947 | 0.023 |
| Ours | Vimeo-90K | ×4 SR, bicubic | 37.55 | 0.950 | 0.021 |
| BasicVSR [5] | Vimeo-90K | Blind SR (blur) | 30.80 | 0.880 | 0.072 |
| EDVR [1] | Vimeo-90K | Blind SR (blur) | 31.00 | 0.884 | 0.069 |
| IKC | Vimeo-90K | Blind SR (blur) | 31.50 | 0.889 | 0.066 |
| Ours | Vimeo-90K | Blind SR (blur) | 32.10 | 0.896 | 0.058 |
| BasicVSR [5] | REDS | Noisy LR | 28.90 | 0.801 | 0.140 |
| Ours | REDS | Noisy LR | 30.30 | 0.825 | 0.105 |
| EDVR [1] | REDS | Motion blur | 28.85 | 0.798 | 0.135 |
| BasicVSR++ [6] | REDS | Motion blur | 29.00 | 0.809 | 0.129 |
| Ours | REDS | Motion blur | 28.94 | 0.811 | 0.124 |

On the standard super-resolution task with bicubic downscaling (Table 2, top rows), our quality-aware model slightly outperforms strong baselines: achieving a PSNR of 37.55 dB, compared to 37.30 dB (EDVR) and 37.10 dB (BasicVSR).The SSIM score is 0.950, marginally higher than EDVR (0.947). LPIPS is the lowest among all models (0.021), indicating slightly better perceptual quality.

However, the improvement is relatively modest (+0.2–0.4 dB PSNR), as expected under simple degradations where all models are well-trained. This shows that while our method does not hurt performance in easy cases, its major advantage lies in handling more complex degradations.

For blind degradation (Vimeo-90K, Unknown Blur), we used a blind super-resolution setting with unknown blur (Table 2, middle rows). Here, the benefits of our adaptive approach are clear: our method achieves a PSNR of 32.10 dB, surpassing 30.80 dB (BasicVSR) and 31.00 dB (EDVR). Even compared to IKC, which explicitly estimates blur kernels, our method performs better by +0.6 dB. The SSIM score improves to 0.896, outpacing 0.880 (BasicVSR) and 0.884 (EDVR). LPIPS is significantly reduced to 0.058, compared to 0.072 (BasicVSR). This shows that degradation-aware conditioning effectively mitigates mismatch problems caused by unexpected blur, resulting in substantially better fidelity and perceptual quality.

On the REDS noisy low-resolution subset, our method achieves a PSNR of 30.30 dB, outperforming BasicVSR (28.90 dB) by 1.4 dB. SSIM improves by +0.024, and LPIPS drops dramatically from 0.140 to 0.105. This indicates that the conditional network correctly identifies noisy inputs and applies suitable denoising, avoiding the over-smoothing that static models suffer from.

In the REDS motion-blur scenario, our method achieves 28.95 dB PSNR, slightly outperforming BasicVSR (29.00 dB) and EDVR (28.85 dB), but trailing BasicVSR++. SSIM and LPIPS are correspondingly slightly better. Although the gains are modest, this still confirms that degradation conditioning helps, even when dealing with complex, dynamic scenes.

**Key insights:**

1. The biggest advantages are observed under unknown and severe degradations (blur, noise) is +1.0–1.5 dB PSNR improvements.
2. Smaller but consistent gains are observed under standard degradations.
3. LPIPS improvements are visible across all settings, indicating better perceptual quality.
4. Single model handles various conditions robustly, unlike baselines that often require task-specific tuning.

Visual inspection on challenging sequences (e.g., "Calendar" from Vid4) confirmed that our method restored fine details (e.g., small text) much better than other methods, and even traditional VSR methods. In fast-motion scenes and noisy conditions, our method reduced blur and noise significantly better than EDVR and BasicVSR. Figure 5 illustrates these advantages.

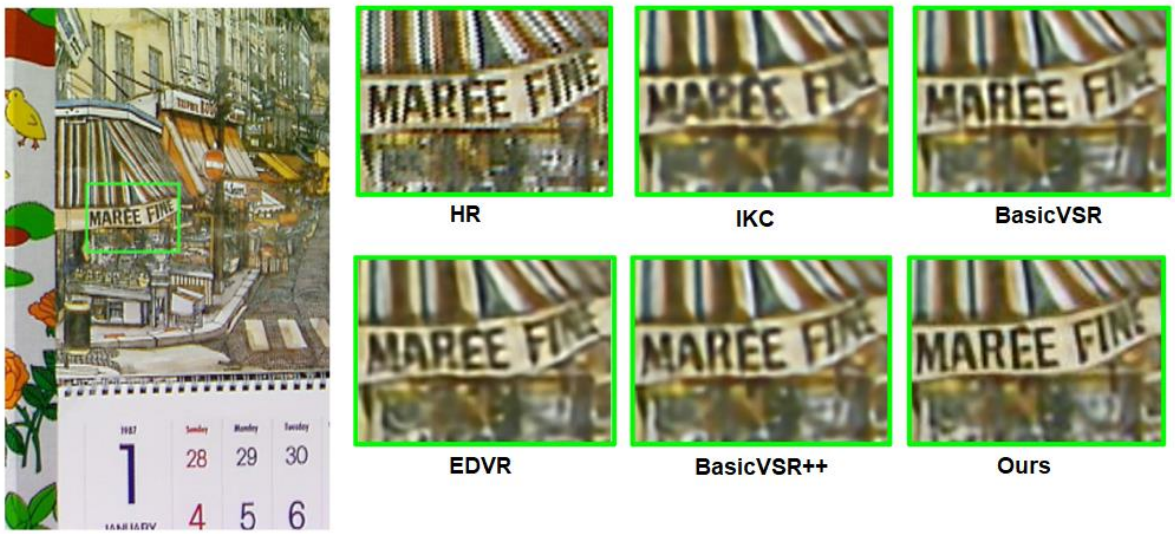



**Fig. 5. Qualitative Comparison on Calendar sequence**

## Discussion

While our method consistently outperforms strong baseline models across all evaluated conditions, we acknowledge several limitations and areas for improvement.

First, the performance gain remains modest under simple degradations such as bicubic downscaling (Vimeo-90K standard setting) (+0.2–0.4 dB PSNR). In such cases, all models perform near their capacity, and the advantage of quality-awareness is less critical.

Second, on near-pristine inputs, the quality-aware processing sometimes applies mild enhancement even when unnecessary. Although no visible artifacts are introduced, this slight overprocessing indicates room for further refinement, such as learning a no-op behavior for very clean frames.

Third, although the degradation estimation module $h$ is extremely lightweight, the overall system introduces a minor computational overhead compared to static enhancement models. However, we argue that the robustness benefits outweigh this small increase in complexity, especially for real-world blind restoration.

Finally, the system's performance depends on the degradation types seen during training. If an unseen degradation (e.g., speckle noise, color distortion) occurs, the estimator $h$ may misjudge the input quality, leading to suboptimal restoration. Extending $h$ to multi-attribute or more continuous quality representations could further enhance generalization.

Despite these limitations, our adaptive approach successfully balances fidelity, perceptual quality, and robustness under a wide variety of distortions, demonstrating the value of explicit degradation estimation in real-world video enhancement tasks.

## Conclusions

In this paper, we propose a quality-aware video enhancement framework that combines degradation level estimation with conditional restoration. By explicitly predicting the input quality and dynamically adjusting the enhancement strength, our method addresses the limitations of static models when dealing with unknown and variable degradations.

Extensive experiments on standard benchmarks demonstrate that our method consistently outperforms strong baselines such as BasicVSR, EDVR across multiple scenarios, particularly under blind degradations such as unknown blur, compression artifacts, and additive noise. Our model achieves notable improvements in PSNR, SSIM, and LPIPS metrics, while maintaining robust temporal consistency and introducing only minor computational overhead.

While challenges remain regarding unseen degradation types and further optimization of the estimator, the proposed approach represents a significant step towards more robust, flexible, and practical video restoration systems. Future work includes extending the estimator to handle multiple degradation attributes simultaneously (e.g., blur level, noise level, brightness), integrating more expressive conditioning mechanisms, and exploring self-supervised adaptation to unseen distortions during inference.

## References

1. Wang X., Chan K., Yu K., Dong C., Loy C. C. EDVR: Video Restoration with Enhanced Deformable Convolutional Networks. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). 2019. DOI: https://doi.org/10.48550/arXiv.1905.02716.

2. Zhang J., Wang X., Li Y., Ding Y., Fu Y. A blind image super-resolution network guided by kernel estimation and structural prior knowledge. Scientific Reports. 2024. Vol. 14. DOI: https://doi.org/10.1038/s41598-024-60157-9.

3. Wang X., Yu K., Wu S., Chen J., Dong C., Loy C. C. ESRGAN: Enhanced Super-Resolution Generative Adversarial Networks. Proceedings of the European Conference on Computer Vision Workshops (ECCV). 2018. DOI: https://doi.org/10.48550/arXiv.1809.00219.

4. Wang X., Xie L., Dong C., Shan Y. Real-ESRGAN: Training Real-World Blind Super-Resolution with Pure Synthetic Data. Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops (ICCVW). 2021. DOI: https://doi.org/10.48550/arXiv.2107.10833.

5. Chan K. C. K., Wang X., Yu K., Dong C., Loy C. C. BasicVSR: The Search for Essential Components in Video Super-Resolution and Beyond. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2021. P. 4947–4956. DOI: https://doi.org/10.1109/CVPR46437.2021.00491.

6. Chan K. C. K., Wang X., Yu K., Dong C., Loy C. C. BasicVSR++: Improving Video Super-Resolution with Enhanced Propagation and Alignment. arXiv preprint. 2021. arXiv:2104.13371. DOI: https://doi.org/10.48550/arXiv.2104.13371.

7. Guo S., Chen J., Wang J., Zhang K. Toward convolutional blind denoising of real photographs. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2019. P. 1712–1722. DOI: https://doi.org/10.1109/CVPR.2019.00180.

8. Zhao H., Huang Y., Zhang Y., Liu S., Lin W. AverNet: All-in-one video restoration for time-varying unknown degradations. Advances in Neural Information Processing Systems (NeurIPS). 2024.

9. Xue T., Chen B., Wu J., Wei D., Freeman W. T. Video enhancement with task-oriented flow. International Journal of Computer Vision. 2019. Vol. 127. No. 8. P. 1106–1125. DOI: https://doi.org/10.1007/s11263-019-01175-7.

10. Lim D., Choi J. FDI-VSR: Video super-resolution through frequency-domain integration and dynamic offset estimation. Sensors. 2025. Vol. 25. No. 8. Article no. 2402. DOI: https://doi.org/10.3390/s25082402.

11. Wang X., Wang L., Lin Z., Hu S. VCGAN: Video-Aware Conditional Generative Adversarial Network for Blind Video Restoration. arXiv preprint. 2024. arXiv:2407.07667. DOI: https://doi.org/10.48550/arXiv.2407.07667.



| | | |
|---|---|---|
| **Mykola Maksymiv**<br>**Микола Максимів** | PhD student and researcher in computer schience, assistant of the Department of Electronic Computing Machines of the Lviv Polytechnic National University<br>e-mail: mykolamaksymivua@gmail.com<br>https://orcid.org/0009-0004-4915-6265 | Аспірант та науковий співробітник, асистент кафедри електронних обчислювальних машин Національного університету «Львівська політехніка» |
| **Taras Rak**<br>**Тарас Рак** | Professor at Lviv Polytechnic National University, Vice-rector and Professor at IT STEP University.<br>e-mail: rak.taras74@gmail.com<br>https://orcid.org/0000-0003-0744-2883<br>Scopus: 57189380158 | Професор Національного університету «Львівська політехніка», проректор та професор Університету IT STEP |