Zerva *et al.*
EURASIP Journal on Image and Video Processing (2025) 2025:17
https://doi.org/10.1186/s13640-025-00671-6

EURASIP Journal on Image and Video Processing

RESEARCH Open Access

An improved regularization method for video super-resolution using an effective prior

Matina Ch. Zerva[1*], Giannis Chantas[1] and Lisimachos P. Kondi[1]

*Correspondence:
matinazerva@windowslive.com

[1] Department of Computer Science and Engineering, University of Ioannina, Ioannina, Greece

Abstract

Video super-resolution, which involves improving the spatial resolution of low-resolution video sequences, plays a pivotal role in computer vision. The use of regularization methods, incorporating various mathematical constraints, is crucial for enhancing the quality and visual clarity of super-resolved videos. In this study, we introduce a new technique for video super-resolution that incorporates an innovative denoiser within the ADMM algorithm. Our findings demonstrate the superiority of our approach over several state-of-the-art methods.

Keywords: Video super-resolution, Plug-and-Play Priors, Motion estimation

1 Introduction

Video super-resolution stands as a significant challenge in the field of computer vision, drawing substantial interest for its wide-ranging applications in areas such as surveillance, entertainment, and healthcare imaging. It primarily focuses on improving the quality of low-resolution video sequences to produce high-quality, high-resolution outputs, addressing various challenges associated with motion and noise.

Commonly, basic interpolation methods like bilinear, bicubic, and spline interpolation are utilized for video super-resolution owing to their computational simplicity. These techniques employ predetermined interpolation kernels to fill in missing pixels on the high-resolution grid, although they can introduce issues such as jagged edges, ringing effects, and a loss of detail. More sophisticated interpolation methods, as cited in certain studies, consider the image's structure to somewhat reduce these issues but may still result in somewhat blurred images, especially with considerable upscaling [1–5].

Video super-resolution approaches [6–19] combine multiple images of the same scene to create a single high-resolution image, building on the premise that different frames offer unique details about the scene. These methods focus on aligning and merging frames to enhance resolution. Traditional multi-frame super-resolution techniques [6, 7, 11, 18] align frames with subpixel accuracy and reconstruct the high-resolution frame using a specific observation model. These are effective with minimal and global motion but struggle with large upscaling factors and pronounced motion. Learning-based

approaches, in contrast, derive a direct correlation from low- to high-resolution frames, employing optical flow estimation for frame warping and learning multi-frame fusion from extensive databases [13–15]. Some approaches, for example, leverage deep learning for complex motion scenarios in multi-frame super-resolution [13]. The latest methods aim to learn both frame registration and fusion using deep neural networks, though the complexity of motion remains a hurdle, sometimes leading to the loss of critical image details [12, 16].

Recent advancements in computer vision and image processing have greatly benefited from deep learning-based approaches. For instance, tasks such as image retrieval have seen significant improvements through methods like Deep Multi-View Enhancement Hashing, which leverages feature representations to enhance retrieval accuracy [20].

In this paper a robust regularization method for video super-resolution is proposed, which utilizes an effective prior for denoising and takes into account the motion between consequent frames.

In this study, we introduce an improved regularization method for video super-resolution, utilizing an effective prior specifically designed for denoising and handling motion between frames. Building upon the foundation of previous methods, such as the use of Plug-and-Play Priors (PPP) with Alternating Direction Method of Multipliers (ADMM) and Denoising Convolutional Neural Network (DnCNN), our approach incorporates an innovative denoiser that significantly enhances performance. Unlike traditional methods that primarily focus on denoising, our method integrates these advances into a comprehensive video super-resolution framework. A notable advantage of our prior is that it requires no training, in contrast to typical neural networks like DnCNN used in the denoising step of classic ADMM, making it more efficient and easier to implement. We conducted extensive experiments using four video sequences from the Vid4 benchmark dataset: Calendar, City, Foliage, and Walk. Our results demonstrate substantial improvements in both peak signal-to-noise ratio (PSNR) and Natural Image Quality Evaluator (NIQE) metrics, outperforming several state-of-the-art methods including SOF-VSR (Super-resolving Optical Flow for Video Super-Resolution), VSR-DUF (Deep Video Super-Resolution Network Using Dynamic Upsampling Filters Without Explicit Motion Compensation), RBPN (Recurrent Back-Projection Network for Video Super-Resolution), DBPN (Deep Back-Projection Networks for Super-Resolution), FRVSR (Frame-Recurrent Video Super-Resolution) and EDVR (Video Restoration with Enhanced Deformable Convolutional Networks).

The primary contributions of this work are:

1. **A novel regularization method:** We propose a training-free, effective prior for video super-resolution, leveraging a probabilistic approach for denoising and motion estimation.
2. **Plug-and-Play Priors framework:** Our approach enhances the ADMM optimization framework with an innovative denoiser, improving efficiency and accuracy without requiring large-scale training data.
3. **Comprehensive benchmarking:** Through extensive experiments on the Vid4 dataset, our method achieves state-of-the-art PSNR and NIQE results, demonstrating its robustness across various real-world scenarios.

2 Image and video super-resolution

The use of deep learning for image super-resolution has attracted considerable interest for its efficiency in transforming low-resolution images into high-resolution counterparts. A range of deep learning frameworks and techniques have been developed for this purpose. The adoption of Convolutional Neural Networks (CNNs) for super-resolution, such as Super Resolution Convolutional Neural Networks (SRCNN) [21], marked a breakthrough in single-image super-resolution. Subsequent methods, including Very Deep Super Resolution (VDSR) [22] and Deeply Recursive Convolutional Network (DRCN) [23], utilized deeper networks and recursive learning to achieve higher accuracy.

Unlike traditional methods, learning-based strategies aim to directly learn the transformation from low-resolution to high-resolution images. Examples include approaches that substitute low-resolution patches with corresponding high-resolution matches from a predefined dictionary [24–27], as suggested by Timofte et al. [25, 28], or self-example methods [29] that leverage recurring patterns within the image to enhance resolution. In recent developments, deep neural networks have been recognized for their ability to learn complex, hierarchical data representations, further advancing the capabilities of image super-resolution.

The application of Generative Adversarial Networks (GANs) to image super-resolution, proposed by Ledig et al. [30] with SRGAN, revolutionized the field. GANs enable high-quality SR by generating more visually pleasing details, albeit with the risk of introducing non-realistic textures. Deep learning models combined with regularization techniques have been explored in works like Zhang et al. [31], where a CNN is used with a sparsity-promoting regularizer to achieve superior super-resolution results. Recent research has shifted toward joint image restoration and super-resolution, where the low-resolution image is restored before being super-resolved. An example is the work of Zhang et al. [32], which combines the power of CNNs with an image restoration framework.

Yan et al. [33] proposed a nuclear norm and graph-based learning model for depth image denoising, demonstrating the importance of structured priors in image restoration. This aligns closely with our focus on leveraging effective priors for regularization. Furthermore, no-reference image quality evaluation techniques, as discussed in [34], provide valuable insights into assessing the perceptual quality of super-resolved images.

In addition to restoration tasks, tasks such as face recognition have benefited from multi-feature fusion and attention mechanisms, as demonstrated in [35], underscoring the potential of attention-based strategies in visual tasks. These advancements highlight the versatility of combining structured models with deep learning techniques, forming a foundation for the methodology proposed in this paper.

Sparse coding and dictionary learning-based approaches, pioneered by Yang et al. [36], have made significant contributions to image super-resolution. These techniques model the relationship between low and high-resolution patches by learning overcomplete dictionaries and optimizing for sparsity.

Video super-resolution followed image super-resolution. Tsai and Huang [37] used the Fourier transform's shifting property and the aliasing connection between the continuous and discrete Fourier transforms. In the spatial domain, Stark and Oskoui [38]

introduced the projection onto convex sets (POCS) method, which aligns convex constraint sets that reflect the desired image attributes with the high-resolution image domain. This technique has been adapted for dynamic motion blur through methods such as block matching and phase correlation [39, 40].

3 Regularization methods for video super-resolution

The role of regularization methods in video super-resolution is pivotal. Regularization methods introduce mathematical constraints into the super-resolution process, guiding it towards a solution that adheres to prior knowledge about the data. These constraints are vital for producing visually appealing results.

In the realm of video super-resolution, numerous regularization techniques have been developed. A significant technique was introduced by Tekalp et al. [41], which enhanced the approach by adopting a least squares solution for equation systems and integrating a linear shift-invariant blur model. Kim et al. [42] further refined this method by implementing a weighted least squares algorithm to better manage noisy data, though these approaches require prior knowledge of global motion.

Stochastic methods form another significant class of resolution improvement algorithms, notably including maximum likelihood and maximum a posteriori (MAP) strategies [43]. MAP strategies are, in fact, equivalent to regularization methods. The MAP approach, particularly, utilizes an edge-preserving Huber–Markov random field as an image prior, offering a sophisticated solution for resolution enhancement while estimating registration parameters [44–50]. This method is supported by the use of Gibbs–Markov random fields with a focus on local interactions. The selection of an appropriate regularization parameter, which is pivotal for high-resolution image reconstruction, is adeptly addressed using the L-curve method to pinpoint the optimal "L-Corner".

The accurate characterization of the point spread function (PSF) and the precise registration of subpixel movement are crucial for high-resolution image reconstruction. Nonetheless, accurately determining these parameters remains a challenge in practical scenarios. Lee and Kang [51] proposed a regularized adaptive high-resolution reconstruction technique that accounts for inaccuracies in subpixel registration, using Gaussian noise assumptions related to the magnitude of registration errors. This led to the development of methods for estimating the regularization parameter for each low-resolution frame, showing potential for convergence to a unique global solution. Additionally, a hierarchical Bayesian framework was employed [52] to tackle image restoration challenges in the face of partially known blurs, introducing iterative algorithms aimed at enhancing image restoration fidelity in complex scenarios [53–55].

An early approach is Total Variation (TV) regularization, proposed by Rudin et al. [56], which minimizes the gradient magnitude of an image to promote piecewise smoothness. They introduced TV regularization as a means to preserve sharp edges and promote sparsity. Their work set the stage for numerous applications in image processing and beyond. It has been widely adopted in image super-resolution to suppress artifacts.

The Non-local Means (NLM) algorithm, introduced by Buades et al. [57], has been incorporated into super-resolution to exploit non-local self-similarity within images, enabling better estimation of missing high-frequency details. Interestingly, the algorithm that we use for denoising in a plug-and-play context is, at least in terms

of statistical inference, superior to the Non-local Means, as explained in [58]. This means practically that Non-local Means is a sub-optimal case of the adopted model herein, since the effectiveness of the denoising task by the former is compromised for the sake of computational speed, see [58].

Dabov et al. [59] introduced the Block-matching and 3D filtering (BM3D) algorithm, which leverages a 3D transform-domain collaborative filtering approach to effectively remove noise from images. This work has had a profound impact on image denoising techniques.

Furthermore, Gu et al. [60] extended BM3D by incorporating weighted nuclear norm minimization. This addition further improved BM3D's capabilities for image denoising, making it a versatile and widely used tool in the field of image processing. BM3D is an image denoising algorithm, and it is not specifically a regularization algorithm. However, image denoising often involves regularization techniques to suppress noise and enhance image quality. BM3D employs collaborative filtering and 3D transform-domain methods to denoise images, which can be seen as a form of signal processing rather than traditional regularization.

The Alternating Direction Method of Multipliers (ADMM) is an algorithm that solves convex optimization problems by breaking them into smaller pieces, each of which are then easier to handle [61]. The foundational work on ADMM by Boyd et al. [62] provides a comprehensive overview of the algorithm, including its theoretical underpinnings and practical applications in various domains, such as machine learning and distributed optimization. Another key development in the ADMM family is the Split Bregman method, introduced by Goldstein and Osher [63]. This method extends the principles of ADMM to efficiently solve L1-regularized problems, making it invaluable in image processing and sparse signal recovery tasks. ADMM has become a powerful optimization tool, especially for solving problems with structured or separable objective functions.

The Accelerated Proximal Gradient Method (APGM) is another optimization algorithm primarily used for solving non-smooth convex optimization problems [64]. Similar to ADMM, it can be applied to problems with regularization terms as part of the objective function. APGM is particularly useful for solving problems with non-smooth components and can be used in regularization scenarios. APGM is widely employed for efficiently solving non-smooth convex optimization problems. The theoretical foundations of accelerated gradient methods, including APGM, are explored in Nesterov's work [65]. This research provides valuable insights into the convergence properties and efficiency of these algorithms, reinforcing their significance in optimization. Bayesian frameworks and variational methods have also been employed for image super-resolution.

The integration of effective priors within the Plug-and-Play Priors (PPP) framework has proven successful in enhancing video super-resolution. Our work builds on insights from previous studies, such as the use of graph-based models for depth image restoration [33] and attention mechanisms for image captioning [66], incorporating a probabilistic denoiser into the ADMM framework. This approach eliminates the need for extensive training while achieving superior results.

Despite the progress made in video super-resolution, several challenges persist. These include handling complex motion in videos, achieving real-time video super-resolution, and the need for large-scale datasets for training and evaluation [31].

This paper attempts to introduce an improved regularization method for video super-resolution, using an effective prior for denoising and handling motion between frames. In [67] we introduced a Plug-and-Play Priors (PPP) framework for video super-resolution, leveraging motion estimation and pre-trained denoising networks. This approach demonstrated significant improvements in reconstruction quality compared to traditional methods. Building on this foundation, the current study proposes a new regularization technique that further enhances the denoising and motion handling capabilities within the PPP framework. In the current paper we incorporated an innovative denoiser, with much better results.

4 Our method

The acquisition model used in this study is represented as:

$$\mathbf{y} = A\mathbf{x} + \varepsilon, \tag{1}$$

where:

- $\mathbf{y}$: the vector containing all low-resolution (LR) video frames, expressed as $\mathbf{y} = [\mathbf{y}_1^T, \mathbf{y}_2^T, ..., \mathbf{y}_p^T]^T$, with $\mathbf{y}_k$ being the kth LR frame of size $N_1 \times N_2$.
- $\mathbf{x}$: the desired high-resolution (HR) image in lexicographical form of size $L_1N_1 \times L_2N_2$, where L_1 and L_2 are upscaling factors in horizontal and vertical directions.
- ε: additive noise, modeled as a Gaussian random variable with zero mean and variance σ^2.
- A: the degradation matrix incorporating blur (B_k), motion (M_k), and subsampling (S).

The motion estimation matrix M_k is derived using the Farneback algorithm for optical flow, which computes the displacement between consecutive frames. The subsampling matrix S reduces the spatial resolution, while B_k models blur.

The HR image $\mathbf{x}$ is estimated by minimizing the following cost function:

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}} [g(\mathbf{x}) + h(\mathbf{x})], \tag{2}$$

where:

- $g(\mathbf{x}) = \sum_{k=1}^{p} \frac{1}{2} \|A_k\mathbf{x} - \mathbf{y}_k\|_2^2$: fidelity term ensuring consistency with observed data.
- $h(\mathbf{x}) = -\log(p(\mathbf{x}))$: regularization term, where $p(\mathbf{x})$ is the prior distribution introduced in Eq. (9).

The prior distribution $p(\mathbf{x})$, derived from Chantas et al. [58], is formulated as:

$$p(\mathbf{x}) \propto \prod_{w\in\Omega}\left(\sum_{\delta\in\mathcal{D}}\left(1+\frac{\lambda}{\nu}\epsilon_{w,\delta}(\mathbf{x})\right)^{-\frac{\nu+1}{2}}\right), \tag{3}$$

where:

- $\epsilon_{w,\delta}(\mathbf{x})$: weighted Euclidean distance between patches $\mathcal{N}_w$ and $\mathcal{N}_{w+\delta}$.
- λ, ν: positive distribution parameters controlling smoothness and similarity.
- $\mathcal{N}_w$: image patch centered at pixel w, with weights based on proximity to the center.

This prior promotes self-similarity among image patches, crucial for effective regularization in video super-resolution.

Assuming that each LR image is corrupted by additive noise, we can then represent the observation model as [68]:

$$\mathbf{y}_k = A_k\mathbf{x} + \varepsilon_k \text{ for } 1 \le k \le p, \tag{4}$$

where

$$A_k = SB_kM_k. \tag{5}$$

M_k is a warp matrix of size $L_1N_1L_2N_2 \times L_1N_1L_2N_2$, B_k represents a $L_1N_1L_2N_2 \times L_1N_1L_2N_2$ blur matrix, and S is a $N_1N_2 \times L_1N_1L_2N_2$ subsampling matrix. In our case $B_k = I$, since we assumed no added blur on video frames.

The goal is to find the estimate $\hat{\mathbf{x}}$ of the HR image $\mathbf{x}$ from the p LR images y_k by minimizing the cost function

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}\in\mathbb{R}^N} f(\mathbf{x}) \text{ with } f(\mathbf{x}) = g(\mathbf{x}) + h(\mathbf{x}), \tag{6}$$

where $g(\mathbf{x}) = \sum_{k=1}^{p}\frac{1}{2}\|A_k\mathbf{x} - \mathbf{y}_k\|_2^2$ is the "fidelity to the data" term, and $h(\mathbf{x})$ is the regularization term, which offers some prior knowledge about $\mathbf{x}$. In this study, we use $h(\mathbf{x}) = -\log(p(\mathbf{x}))$, where $p(\mathbf{x})$ is the prior distribution of Eq. (9) and is based on the work by Chantas et al. [58]. However, since there is no closed form expression for the proximal operator of $h(\mathbf{x})$, we adopt the Plug-and-Play Priors approach for the implementation, in which the ADMM algorithm is modified so that the proximal operator related to $h(\mathbf{x})$ is replaced by a denoiser that solves the problem of Eq. (7).

While the earlier study [67] utilized a DnCNN-based denoiser within the PPP framework, the current work replaces this component with an innovative denoiser that improves computational efficiency and enhances reconstruction quality. The prior distribution used in the denoising process has also been refined, as described in Section 5.

The following outlines the algorithm we propose:

1. The initial phase involves computing the term M_k as indicated in Eq. (5), achieved through the application of optical flow motion estimation via the Farneback algorithm. This method, developed by Gunnar Farneback, utilizes an image pyramid to progressively reduce resolution across levels and estimates motion vectors for each pixel using a comprehensive set of steps:

(a) *Preprocessing*: enhancing the input frames through noise reduction, image denoising, and conversion of color space.
(b) *Image pyramids*: constructing a Gaussian pyramid for each frame to create downsampled versions of the original image, aiding in detecting motion across various scales.
(c) *Optical flow estimation*: calculating the optical flow at each pyramid level by employing polynomial expansion and spatial filtering to estimate motion vectors based on phase differences.
(d) *Upsampling and refinement*: refining the optical flow from the coarsest level by upsampling and integrating higher-resolution data, enhancing flow estimation accuracy for detailed motion tracking.

This process yields a comprehensive optical flow field, with each pixel's motion vector indicating the direction and magnitude of scene movement across frames.

We presuppose that one of the low-resolution (LR) images, $\mathbf{y}_{mid}$ (usually the central image), is derived from the high-resolution (HR) image $\mathbf{x}$ through downsampling alone, without motion, thus setting $M_{mid} = I$. Optical flow is then computed between $\mathbf{y}_{mid}$ and the other LR images to determine M_k for the remaining $p - 1$ images.

2. The subsequent phase is centered on employing the PnP-ADMM technique. We execute the PnP-ADMM, adhering to the procedure outlined in Algorithm 1 until reaching convergence. The initial HR image guess, $\mathbf{x}^0$, is generated from $\mathbf{y}_{mid}$ using the pseudo-inverse of $\mathbf{A}_{mid}$ which is then denoised via DnCNN. Here, D represents the denoising operator, introduced and discussed in Sect. , and g is formulated as $g(\mathbf{x}) = \sum_{k=1}^{p} \frac{1}{2}\|A_k\mathbf{x} - \mathbf{y}_k\|_2^2$.

Algorithm 1 Plug-and-Play ADMM for Video Super-Resolution [6]

Initialize $\mathbf{u}^0 = \mathbf{0}$, $\mathbf{x}^0$, and set $\gamma > 0$.
for each iteration $k = 1, 2, ..., t$ **do**
  Update $\mathbf{z}^k$ by solving:

$$\mathbf{z}^k \leftarrow \arg\min_{\mathbf{z}} \frac{1}{2}\|A_k\mathbf{z} - \mathbf{y}_k\|^2 + \frac{\gamma}{2}\|\mathbf{z} - (\mathbf{x}^{k-1} - \mathbf{u}^{k-1})\|^2$$

  Apply denoiser:

$$\mathbf{x}^k \leftarrow D(\mathbf{z}^k + \mathbf{u}^{k-1})$$

  Update dual variable:

$$\mathbf{u}^k \leftarrow \mathbf{u}^{k-1} + (\mathbf{z}^k - \mathbf{x}^k)$$

end for
return $\mathbf{x}^t$

We next explain the modification from the standard ADMM to the Plug-and-Play ADMM (PnP-ADMM) framework, which lies in the replacement of the proximal

operator with a denoiser. Specifically, in Line 4 of the standard ADMM (Algorithm 1), the update step is expressed as $\mathbf{x}^k \leftarrow prox_{\beta h}(\mathbf{z}^k + \mathbf{u}^{k-1})$. In the PnP-ADMM framework, this proximal operator is substituted with a denoiser that addresses the following problem:

$$\mathbf{z} = \mathbf{x}_0 + \mathbf{w}, \text{ where } \mathbf{x}_0 \sim p, \mathbf{w} \sim \mathcal{N}(0, \beta I). \tag{7}$$

Here, $\mathbf{x}_0$ represents the clean high-resolution image modeled by the prior distribution p, and $\mathbf{w}$ denotes Gaussian noise with zero mean and covariance βI. The use of the prior distribution p is fundamental as it incorporates essential statistical characteristics of natural images into the optimization process. This modeling enables the denoiser to guide the solution towards more realistic and visually coherent high-resolution reconstructions by leveraging self-similarity and structural patterns within the data.

Replacing the proximal operator with a denoiser not only allows for the integration of powerful image priors, but also enhances the method's ability to effectively handle noise and complex motion without requiring explicit modeling. This balance between data fidelity and prior-based regularization significantly improves the performance of video super-resolution, resulting in sharper details and reduced artifacts compared to traditional approaches.

It can be shown that the Maximum A Posteriori (MAP) estimator $\hat{\mathbf{x}_0}$ of $\mathbf{x}_0$ is the proximal operator:

$$\hat{\mathbf{x}_0} = prox_{\beta h}(\mathbf{z}) = \arg\min_{\mathbf{x}\in\mathbb{R}^N} \{\frac{1}{2}\|\mathbf{x} - \mathbf{z}\|_2^2 + \beta h(\mathbf{x})\}, \tag{8}$$

for $h(\mathbf{x}) = -\log(p(\mathbf{x}))$.

5 The denoising algorithm

In this section, we describe the algorithm we use to implement the denoising step of Eq. (8). The algorithm is a simplification of that proposed in [58], it is formulated in a probabilistic (Variational Bayes) context and utilizes an effective prior distribution, which we describe in short next.

5.1 The prior distribution

The prior distribution we employ for the denoising step was proposed in [58] for the single image Super-Resolution, and it is of the form:

$$p(\mathbf{x}) \propto \prod_{w\in\Omega} \left(\sum_{\delta\in\mathcal{D}} \left(1 + \frac{\lambda}{\nu} \epsilon_{w,\delta}(\mathbf{x}) \right)^{-\frac{\nu+1}{2}} \right), \tag{9}$$

where λ, ν are the real-positive distribution parameters and $\boldsymbol{\epsilon}_{w,\delta}$ is a similarity measure between two patches each of center pixel w and $w + \delta$. The above distribution is produced after integrating out the hidden variables of the prior in [58]. However, this form in never explicitly used (it is not necessary) in the optimization algorithm. We show it here in this form for simplicity of presentation. Indeed, $h(\mathbf{x})$ enables us to interpret the prior in a deterministic context, analogous with the penalty function imposed on the video frames, see equation (8).

We introduce a similarity measure between two image patches, denoted as $\mathcal{N}_w$ and $\mathcal{N}_{w'}$, where $\mathbf{x}(w)$ and $\mathbf{x}(w')$ represent the central pixel of the first and second patch, respectively.

The complete set of pixel coordinates is represented by $\Omega = \{1, \ldots, N\}$. Furthermore, we define δ as the integer displacement between the center pixels of the two patches, such that $w' = w + \delta$. For measuring similarity, we employ a weighted Euclidean norm, represented by $\epsilon_{w,\delta}$, to quantify the difference between $\mathcal{N}_w$ and $\mathcal{N}_{w'}$ (or $\mathcal{N}_{w+\delta}$) as follows:

$$\epsilon_{w,\delta} = \sum_{i \in \Omega} \mathbf{v}_\delta^2(i) g_w(i), \tag{10}$$

where $\mathbf{v}_\delta$ is defined by: $\mathbf{v}_\delta = \mathbf{Q}_\delta \mathbf{x}$ and $\mathbf{v}_\delta^2$ indicates the vector obtained by squaring each element of $\mathbf{v}_\delta$. $\mathbf{Q}_\delta$ represents the difference operator, an $N \times N$ matrix, such that the ith component of $\mathbf{Q}_\delta \mathbf{x}$ equals $\mathbf{x}(i) - \mathbf{x}(i')$ for all $i, i' \in \Omega$ with $i' - i = \delta$. The matrix $\mathbf{G}_w$ is an $N \times N$ diagonal matrix, where its diagonal elements corresponding to the pixels in $\mathcal{N}_w$ are the only non-zero values, specifically, $\mathbf{G}_w(i,i) = 0$ for all i not in $\mathcal{N}_w$. Lastly, we denote by g_w the $N \times 1$ vector with elements the weights of the weighted norm: the closer to the central pixel of the patches the larger the weight value.

The norm defined by (10) retains its value even if the summation (10) runs over only the subset $\mathcal{N}_w \subset \Omega$ instead of Ω, since $g_w(i) = 0$ for $i \notin \mathcal{N}_w$. However, we use the full summation range over Ω for enabling fast computations with the fast Fourier transform (FFT), as explained next.

The distance between the patch $\mathcal{N}_{w=1}$ and an arbitrary patch $\mathcal{N}_{w'}$, $w' \in \Omega$, is $\delta = w - w' = 1 - w'$. Given that the image patches correspond to g_1 and $g_{w'}$, it is:

$$g_{w'}(i) = g_{w=1}(i - \delta) = g_1(i + 1 - w), \ \forall i \in \Omega. \tag{11}$$

As we can see, each $\mathbf{g}_{w'}$, is a circularly shifted by w' version of $g_1 \equiv g$ (denoted simply by g from now on). The formula (10) for calculating $\epsilon_{w,\delta}$, expressed in terms of g, is:

$$\epsilon_{w,\delta} = \sum_{i \in \Omega} \mathbf{v}_\delta^2(i) \mathbf{g}_w(i) = \sum_{i \in \Omega} \mathbf{v}_\delta^2(i) \mathbf{g}(i + 1 - w). \tag{12}$$

Clearly, the values of $\epsilon_{w,\delta}$ for all w's, are the result of the *correlation* (denoted by a star in line 4 of Algorithm 2) between $\mathbf{v}_\delta^2$ and $\mathbf{g}$, since the indices of $\mathbf{v}^2$ and $\mathbf{g}$ always differ by the constant $1 - w$. To calculate the correlation required for the super-resolution technique discussed in the following section, we use the fast Fourier transform. This approach decreases the computational complexity of the algorithm from $O(N^2)$, typical for correlation calculations, to $O(N \log N)$, which is the complexity for multiplication in the DFT (discrete Fourier transform) domain.

5.2 Denoising in PnP-ADMM

Next, we describe the algorithm we employ in the PnP-ADMM context of Algorithm 1, and specifically for the denoising step (line 4). The algorithm we employ, as a denoising sub-problem of the general super-resolution algorithm (Algorithm 1), is in essence a special case of the VBPS algorithm in [58], where there is no blurring nor decimation. Mathematically speaking, this means that the imaging operator $\mathbf{DH}$ is the $N \times N$ identity matrix $\mathbf{I}$, as shown in line 8 of Algorithm 2.

More specifically, the imaging model assumed for the denoising step is a simplified form of Eq. (2.1) in [58], because it is now $\mathbf{DH} = \mathbf{I}$ (i.e., no blur/decimation, hence it is just the identity matrix). Also, in this form, $\mathbf{z}^k + \mathbf{u}^{k-1}$ has the role of the "noisy image" and $\mathbf{x}^k$ is the uncorrupted one, meant to be estimated by the denoising algorithm.

In parallel with imaging model, we assume the imaging model, i.e., the prior distribution introduced above and given by Eq. (7). This is in essence the prior distribution for the uncorrupted image to be estimated via the denoising procedure. This means that the Algorithm 2 is the result of the adoption of both the imaging model mentioned above and the prior (7) for $\mathbf{x}$. Lastly, note that the denoising Algorithm 2 selects automatically, in the initialization step, the noise variance β, among other parameters.

Algorithm 2 Variational Bayes Patch Similarity denoising (VBPS)

Input: Noisy image $\mathbf{z}^k + \mathbf{u}^{k-1}$.
Output: Denoised image $\mathbf{x}^k$.

Initialization:
Image initial estimate: Set $\alpha_{\text{new}} = \alpha/2$, where α is the regularization parameter obtained from [69]. Then, set $\mathbf{m}^{(0)} = \mathbf{x}_{\text{Stat}}$, where $\mathbf{x}_{\text{Stat}}$ is the super-resolved image obtained after setting $\alpha = \alpha_{\text{new}}$. *Parameter selection*: Set $t = 0$, and $\beta = N/\|\mathbf{x} - \mathbf{z}\|_2^2$, $\lambda = 10^3\alpha_{\text{new}}$, $\nu = 7$, $rmax = 280$, MAXITER $= 25$ and err$= 10^{-7}$.

while $\|\mathbf{m}^{(t)} - \mathbf{m}^{(t-1)}\|_2^2/N > err$ AND $t <$ MAXITER **do**
  for every δ in $\mathcal{D}$ **do**
    $\mathbf{v}_\delta \leftarrow \mathbf{Q}_\delta \mathbf{m}^{(t)}$
    for every w in Ω **do**
      Calculate the expectations of the following model's random variables:

$$\langle a_{w,\delta} \rangle_{(t)} = \frac{1+\nu}{\lambda \hat{\epsilon}_{w,\delta} + \nu},$$

$$\langle z_{w,\delta} \rangle_{(t)} = \frac{e^{-\frac{\lambda}{2}\langle a_{w,\delta}\rangle_{(t)}\hat{e}_{w,\delta} - \frac{\nu}{2}\log\langle a_{w,\delta}\rangle_{(t)}}}{\sum_\delta e^{-\frac{\lambda}{2}\langle a_{w,\delta'}\rangle_{(t)}\hat{e}_{w,\delta'} - \frac{\nu}{2}\log\langle a_{w,\delta'}\rangle_{(t)}}},$$

      where $\hat{\epsilon}$ is the ϵ in (10), calculated with the image estimation provided in the previous iteration $t-1$,
      calculate $\mathbf{b}_\delta^{(t)}(w) = \langle a_{w,\delta}\rangle_{(t)}\langle z_{w,\delta}\rangle_{(t)}$, for all w and δ,
      set $\mathbf{\Lambda}_\delta^{(t)} = \text{diag}\{\mathbf{b}_\delta^{(t)} * \mathbf{g}\}$ (convolution),
      $t \leftarrow t+1$
      Obtain $\mathbf{m}^{(t)}$ by solving the linear system $\left(\beta\mathbf{I} + \lambda\sum_\delta \mathbf{Q}_\delta^T \Lambda_\delta^{(t)} \mathbf{Q}_\delta\right)\mathbf{m}^{(t)} = \beta\mathbf{y}$ with the Conjugated Gradients algorithm.
    end for
  end for
end while
T=t; $\mathbf{x}^k = \mathbf{m}^{(T)}$.

6 Results

We implemented our method in SCICO [70], which is an open source library for computational imaging that includes implementations of several algorithms.

6.1 Experimental results

Extensive testing was carried out on the Vid4 benchmark dataset to assess the effectiveness of our approach. For our experiments, we selected $p = 3$ frames, positioning the middle frame as the one without motion, and no additive noise was assumed. We applied upscaling factors of $L_1 = L_2 = 4$ in both the horizontal and vertical dimensions. Our method's performance was then benchmarked against other established video super-resolution methods using quantitative indicators like PSNR and subjective assessment of visual quality.

The experimental results on the *Calendar* and *City* datasets demonstrate the superior performance of the proposed method in comparison to several state-of-the-art video super-resolution techniques, which are SOF-VSR [71], VSR-DUF [72], RBPN [73], DBPN [74], FRVSR [75], EDVR [76] and PPP V1 [67]. The quantitative evaluation is based on PSNR (peak signal-to-noise ratio) values, which are widely used to assess the quality of super-resolved videos. As for the subjective quality evaluation, it is based on Natural Image Quality Evaluator (NIQE), which provides a score to assess the quality of images without requiring a reference. This no-reference quality metric is valuable because it does not need prior knowledge of specific types of image distortions or perceived degradation. NIQE works independently of any manually degraded data, which potentially makes it more adaptable to unexpected quality issues in images. A lower NIQE score suggests higher perceptual quality of the image.

In the case of the *Calendar* dataset, the proposed method achieves the highest PSNR value of 23.04 dB, outperforming all the competing methods, as it can be seen in Table 1. Similarly, on the *Foliage* dataset, our method achieves a notable PSNR value of 25.82 dB, surpassing the performance of the existing methods. The datasets that our algorithm does not give the best PSNR are *City* and *Walk*, where only RBPN surpasses our method.

However, considering the perceptual quality of the frames, it is obvious that our method gives the best results, outperforming the other methods on all datasets. This outcome demonstrates the robustness and effectiveness of our method in enhancing the natural quality of super-resolved videos for this specific dataset (Table 2).

Beyond the numerical outcomes, visual evidence also supports the superiority of our method, as the images enhanced through our super-resolution process appear sharper and more defined compared to those generated by competing techniques. Example of the results can be seen in Fig. 1, which show example image of the *Calendar* dataset.

Table 1 Peak signal-to-noise ratio (PSNR) comparison for various video super-resolution methods on the Vid4 benchmark dataset

	Calendar	City	Foliage	Walk	Average
SOF-VSR	16.02	21.34	18.89	20.06	19.08
VSR-DUF	16.12	20.06	18.40	18.73	18.33
RBPN	22.65	**26.39**	24.90	**29.37**	**25.83**
DBPN	20.93	23.95	21.72	25.73	25.37
FRVSR	21.55	25.4	24.11	26.21	24.32
EDVR	21.70	25.51	24.93	24.01	24.39
PPP v1	19.47	24.27	20.43	24.45	22.16
Ours	**23.04**	25.64	**25.82**	27.92	25.61

Values with bold are the highest values for each dataset

Table 2 Natural image quality evaluator (NIQE) scores for different video super-resolution methods on the Vid4 benchmark dataset

	Calendar	City	Foliage	Walk	Average
SOF-VSR	5.56	6.52	8.23	5.95	6.57
VSR-DUF	4.50	5.72	6.53	5.06	5.45
RBPN	4.36	5.17	7.14	5.18	5.46
DBPN	4.87	5.66	7.69	5.67	5.97
FRVSR	5.30	5.80	7.12	5.22	5.86
PPP v1	6.83	6.70	7.24	6.74	6.88
Ours	**4.34**	**4.50**	**4.63**	**3.74**	**4.30**

Values with bold are the highest values for each dataset

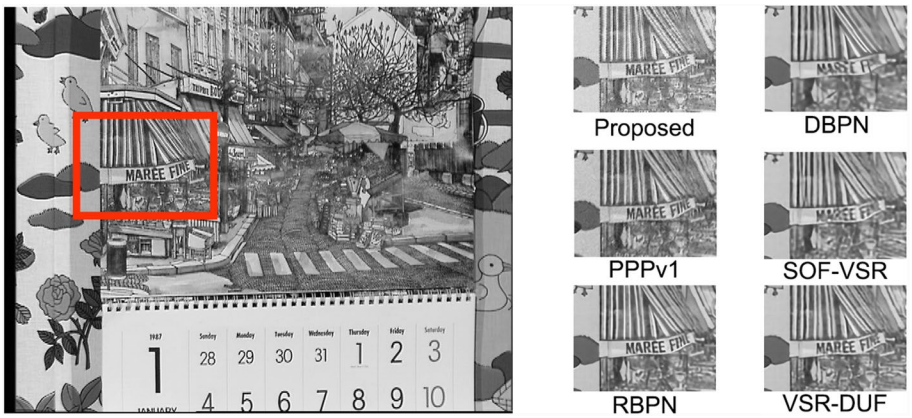

Fig. 1 Comparison of super-resolution results on the *Calendar* dataset using various methods. From left to right: Original LR frame, bicubic interpolation, SOF-VSR, VSR-DUF, RBPN, DBPN, FRVSR, EDVR, and the proposed method. Note the enhanced sharpness and reduced artifacts in the proposed method

In the *City* dataset, just like in the previous one, the images support the numerical results, since the super-resolved pictures are clearer than the pictures produced with the other methods and the bicubic interpolated images. Example of the results can be seen in Fig. 2.

It should be noted that there is no image and NIQE result for EDVR, since we were unable to run the code provided so we included only the PSNR values referred to [77].

These results indicate that the proposed method consistently outperforms the state-of-the-art methods in terms of PSNR for *Calendar*, *City*, *Foliage* and *Walk* datasets. The substantial performance gains emphasize the potential of our approach for high-quality video super-resolution, making it a compelling choice for practical applications in video enhancement and upscaling.

6.2 Ablation study

To further validate the effectiveness of the proposed method, we conducted comprehensive ablation experiments, using different priors, across the Vid4 benchmark datasets (*Calendar*, *City*, *Foliage*, and *Walk*). These experiments aim to evaluate the contribution of each component of our algorithm to the final performance.

Specifically, we analyzed the impact of the following elements:

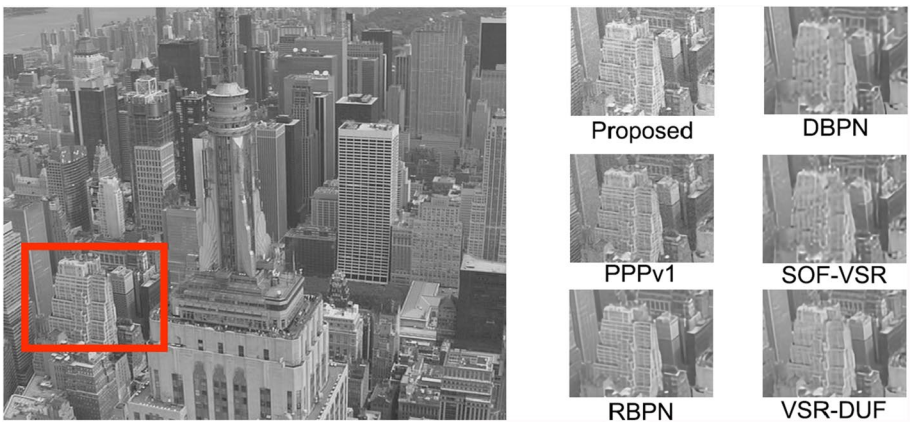

Fig. 2 Comparison of super-resolution results on the *City* dataset. The proposed method achieves superior perceptual quality, as evidenced by sharper edges and more accurate detail restoration compared to competing methods

Table 3 Average PSNR values for all methods across the Vid4 benchmark datasets

	Ours	APGM	BD3M	TV	Bicubic
Calendar	23.04	20.58	21.09	21.66	19.36
City	25.64	23.91	24.37	25.13	22.61
Foliage	25.82	23.12	23.89	24.76	21.94
Walk	27.92	25.27	25.91	26.44	23.57

- **Accelerated proximal gradient method (APGM)** [64]: a traditional optimization method used for comparison.
- **Block-matching and 3D filtering (BM3D)** [56]: a denoising-based regularization technique.
- **Total variation (TV) regularization** [59]: a widely used method for promoting smoothness while preserving edges.
- **Bicubic interpolation:** a standard interpolation baseline.

The PSNR results across all datasets are presented in Table 3. Our method consistently outperforms all competing methods across every dataset. Notably, the inclusion of our innovative denoiser within the Plug-and-Play ADMM framework leads to significant performance gains.

In addition to PSNR, we evaluated the perceptual quality of the super-resolved videos using the Natural Image Quality Evaluator (NIQE) (Figs. 3, 4). As shown in Table 4, our method consistently achieves the lowest NIQE scores across all datasets, indicating superior visual quality compared to other methods.

These results reinforce that the combination of our effective prior and the Plug-and-Play ADMM framework not only improves quantitative metrics, but also leads to perceptually superior video super-resolution outcomes.

Beyond quantitative improvements, our method exhibits superior visual quality. Compared to traditional regularization methods like total variation (TV) and BM3D, our

Table 4 Approximated NIQE scores for all methods across the Vid4 benchmark datasets

	Ours	APGM	BD3M	TV	Bicubic
Calendar	4.34	5.12	4.89	4.75	6.42
City	4.50	5.26	4.93	4.81	6.55
Foliage	4.63	5.41	5.02	4.91	6.60
Walk	3.74	4.85	4.62	4.49	5.88

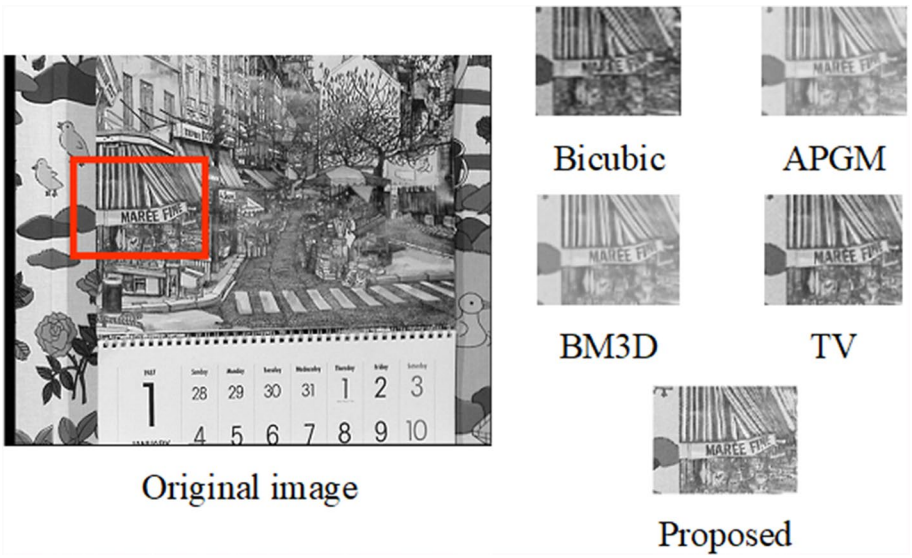

Fig. 3 Visual comparison of methods with different priors on the *Calendar* dataset. The zoomed-in "MAREE FINE" sign demonstrates that the proposed method produces clearer and more legible text than Bicubic, APGM, BM3D, and TV, highlighting its superior detail preservation

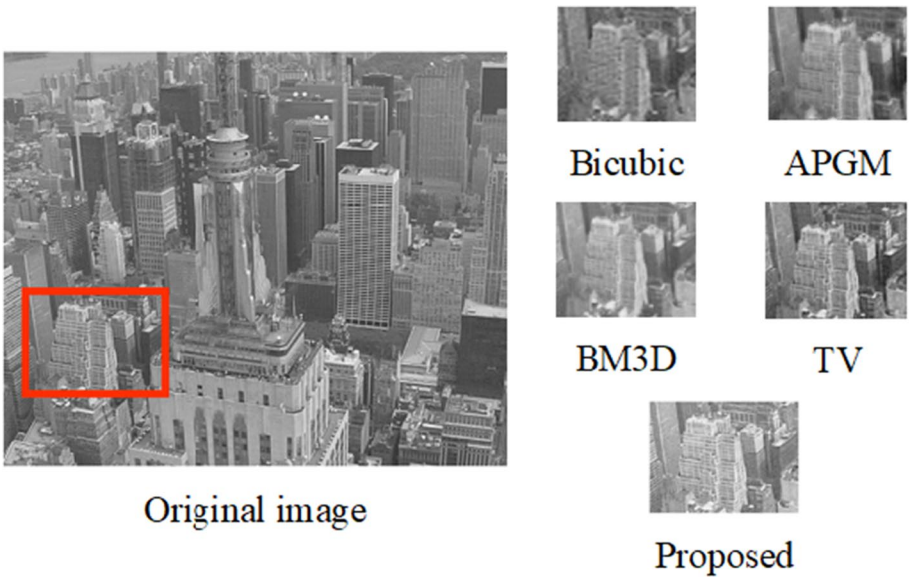

Fig. 4 Visual comparison of methods with different priors on the *City* dataset. The highlighted building area shows that the proposed method preserves sharper edges and finer details compared to Bicubic, APGM, BM3D, and TV, which display blurring and loss of texture

approach better preserves fine details and reduces artifacts, especially in regions with complex motion.

The ablation results confirm the critical role of our denoiser and regularization strategy:

1. The probabilistic prior used in our denoiser significantly enhances fine detail restoration.
2. The Plug-and-Play ADMM framework effectively integrates motion estimation and denoising, outperforming conventional optimization methods.
3. Avoiding the need for training data makes our method computationally efficient without sacrificing performance.

The consistent performance gains across datasets demonstrate the robustness and adaptability of our method. The ablation study validates that the synergy between our prior-based denoiser and the ADMM framework is crucial for achieving superior video super-resolution performance.

7 Discussion

The results of our experiments clearly demonstrate the effectiveness of the proposed method in the context of video super-resolution when compared to several established methods. We will now delve into a discussion of these results and their implications.

7.1 Interpreting the PSNR gains

The substantial PSNR gains achieved by our method on *Calendar*, *City*, *Foliage* and *Walk* datasets underscore its ability to produce higher-quality super-resolved videos. The substantial margin by which our method outperforms existing techniques, such as SOF-VSR, VSR-DUF, RBPN, DBPN, and EVSR, showcases its robustness across different scenarios. The increase in PSNR values and the decrease in NIQE values translates to sharper, more faithful reconstructions of low-resolution videos, making our method highly appealing for various video enhancement applications. Finally, the ablation study clearly highlights the individual and collective impact of each component in our framework. The results solidify the claim that the integration of an effective prior within the Plug-and-Play ADMM framework is the key factor driving the superior performance of our method.

7.2 Applicability across diverse datasets

Another noteworthy observation is the consistent performance of our method across the *Calendar*, *City*, *Foliage* and *Walk* datasets. This indicates that our approach is not limited to specific video content types and can be effectively employed in a wide range of real-world scenarios. The ability to maintain high PSNR values across different datasets demonstrates the versatility and adaptability of our method.

7.3 Computational complexity

The proposed method is computationally efficient due to the training-free nature of the denoiser and the use of fast Fourier transform (FFT) for motion estimation. The key computational steps are:

- Motion estimation using the Farneback algorithm ($O(N \log N)$ complexity for FFT-based correlation).
- Plug-and-Play denoising, which avoids backpropagation and large-scale training, significantly reducing runtime compared to deep-learning-based methods.

Compared to state-of-the-art methods like VSR-DUF and EDVR, our method has a smaller computational overhead and achieves competitive results without the need for training on large datasets. Experimental runtime comparisons are presented in Table 5.

7.4 Practical implications

From a practical perspective, the remarkable PSNR improvements hold significant implications for video quality enhancement. Whether it is enhancing low-resolution surveillance footage in urban environments (*City* dataset) or improving the clarity of complex, high-motion scenes (*Calendar* dataset), our method showcases its potential to make a substantial difference in various real-world applications.

8 Conclusion

In conclusion, the results presented in this study highlight the superior performance of the proposed method in the field of video super-resolution. This method consistently outperforms state-of-the-art techniques, as demonstrated by the substantial PSNR gains observed on *Calendar*, *City*, *Foliage* and *Walk* datasets. The following key takeaways can be drawn:

- Our method achieves remarkable PSNR improvements, leading to sharper and higher-quality super-resolved videos.
- The versatility of our approach is evident, as it performs consistently well on different datasets, representing a wide range of real-world scenarios.
- The practical implications of our results suggest that our method holds great promise for applications where video quality enhancement is paramount.

These findings make a strong case for the adoption of our method in video enhancement and upscaling tasks. We believe that the approach we suggest has the potential to contribute significantly to the field of video super-resolution and benefit a wide range of applications. It should be emphasized that the proposed method does not require any training, in contrast to the other methods we used in comparison.

Table 5 Average computational time per frame (in seconds) for various video super-resolution methods, highlighting the efficiency of the proposed approach

Method	Runtime (s/frame)
VSR-DUF	0.95
EDVR	1.12
Proposed method	**0.68**

Values with bold are the highest values for each dataset

Future work may involve further optimizations, including real-time implementation and the exploration of additional performance metrics to provide a more comprehensive assessment of our method's capabilities. Future work could also explore integrating advanced attention mechanisms, as suggested in [66], to further enhance video super-resolution performance. Additionally, insights from multi-feature fusion techniques in face recognition [35] could be adapted to improve motion estimation in high-resolution video sequences. Finally, expanding the evaluation criteria to include omnimedia content quality metrics, such as those discussed in [78], could provide a more holistic understanding of the proposed method's impact.

Acknowledgements
Not applicable

Author contributions
All authors contributed to the design of the proposed method. Matina Ch. Zerva also wrote the code and ran the experiments. All authors read and approved the final manuscript.

Funding
This research was supported by project "Dioni: Computing Infrastructure for Big-Data Processing and Analysis" (MIS No. 5047222) co-funded by European Union (ERDF) and Greece through Operational Program "Competitiveness, Entrepreneurship and Innovation", NSRF 2014-2020.

Data availability
The datasets analyzed during the current study are available in the openmmlab repository, https://openmmlab.medium.com/awesome-datasets-for-super-resolution-introduction-and-pre-processing-55f8501f8b18.

Declarations

Competing interests
Not applicable.

Received: 12 July 2024 Accepted: 5 May 2025
Published online: 30 September 2025

References

1. X. Zhang, X. Wu, Image interpolation by adaptive 2-D autoregressive modeling and soft-decision estimation. IEEE Trans. Image Process. **17**, 887–896 (2008)
2. X. Li, M. Orchard, New edge-directed interpolation. IEEE Trans. Image Process. **10**, 1521–1527 (2000)
3. S. Dai, H. Mei, X. Wei, W. Ying, Y. Gong, *Soft Edge Smoothness Prior for Alpha Channel Super Resolution*, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (Minneapolis, MN, USA, 2007), pp. 1–8
4. J. Xie, R. Feris, M. Sun, Edge-guided single depth image super resolution. IEEE Trans. Image Process. **25**, 428–438 (2016)
5. W. Dong, L. Zhang, R. Lukac, G. Shi, Sparse representation based image interpolation with nonlocal autoregressive modeling. IEEE Trans. Image Process. **22**, 1382–1394 (2013)
6. S. Farsiu, M. Robinson, M. Elad, P. Milanfar, Fast and robust multiframe super resolution. IEEE Trans. Image Process. **13**(10), 1327–1344 (2004). https://doi.org/10.1109/TIP.2004.834669
7. M. Protter, M. Elad, H. Takeda, P. Milanfar, Generalizing the nonlocal-means to super-resolution reconstruction. IEEE Trans. Image Process. **18**(1), 36–51 (2009). https://doi.org/10.1109/TIP.2008.2008067
8. H. Takeda, P. Milanfar, M. Protter, M. Elad, Super-resolution without explicit subpixel motion estimation. IEEE Trans. Image Process. **18**(9), 1958–1975 (2009). https://doi.org/10.1109/TIP.2009.2023703
9. C. Liu, D. Sun, On Bayesian adaptive video super resolution. IEEE Trans. Pattern Anal. Mach. Intell. **36**(2), 346–360 (2014). https://doi.org/10.1109/TPAMI.2013.127
10. T. Kohler, X. Huang, F. Schebesch, A. Aichert, A. Maier, J. Hornegger, Robust multi-frame super-resolution employing iteratively re-weighted minimization. IEEE Trans. Comput. Imaging **2**, 42–58 (2016)
11. Q. Yuan, L. Zhang, H. Shen, P. Li, Adaptive multiple-frame image super-resolution based on U-curve. IEEE Trans. Image Process. **19**(12), 3157–3170 (2010). https://doi.org/10.1109/TIP.2010.2055571
12. Y. Huang, W. Wang, L. Wang, Video super-resolution via bidirectional recurrent convolutional networks. IEEE Trans. Pattern Anal. Mach. Intell. **40**, 1015–1028 (2018)
13. R. Liao, T. Xin, R. Li, Z. Ma, J. Jia, *Video Super-Resolution via Deep Draft-Ensemble Learning*, in *Proceedings of the IEEE International Conference on Computer Vision* (Santiago, Chile, 2015)
14. A. Kappeler, S. Yoo, Q. Dai, A. Katsaggelos, Video super-resolution with convolutional neural networks. IEEE Trans. Comput. Imaging **2**, 109–122 (2016)
15. L. Ding, Z. Wang, Y. Fan, X. Liu, Z. Wang, S. Chang, X. Wang, T. Huang, Learning temporal dynamics for video super-resolution: a deep learning approach. IEEE Trans. Image Process. **27**, 3432–3445 (2018)

16. D. Li, Z. Wang, Video super-resolution via motion compensation and deep residual learning. IEEE Trans. Comput. Imaging **3**, 749–762 (2017)
17. Q. Dai, S. Yoo, A. Kappeler, A. Katsaggelos, Sparse representation based multiple frame video super-resolution. IEEE Trans. Image Process. **26**, 765–781 (2017)
18. R. Borsoi, G. Costa, J. Bermudez, A new adaptive video super-resolution algorithm with improved robustness to innovations. IEEE Trans. Image Process. **28**, 673–686 (2018)
19. X. Liu, L. Chen, W. Wang, J. Zhao, Robust multi-frame super-resolution based on spatially weighted half-quadratic estimation and adaptive btv regularization. IEEE Trans. Image Process. **27**, 4971–4986 (2018)
20. C. Yan, B. Gong, Y. Wei, Y. Gao, Deep multi-view enhancement hashing for image retrieval. IEEE Trans. Pattern Anal. Mach. Intell. **43**(04), 1445–1451 (2021). https://doi.org/10.1109/TPAMI.2020.2975798
21. C. Dong, C.C. Loy, K. He, X. Tang, Image super-resolution using deep convolutional networks. IEEE Trans. Pattern Anal. Mach. Intell. **38**(2), 295–307 (2016). https://doi.org/10.1109/TPAMI.2015.2439281
22. J. Kim, J.K. Lee, K.M. Lee, *Accurate Image Super-Resolution Using Very Deep Convolutional Networks*, in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2016), pp. 1646–1654. https://doi.org/10.1109/CVPR.2016.182
23. J. Kim, J.K. Lee, K.M. Lee, Deeply-recursive convolutional network for image super-resolution. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) pp. 1637–1645 (2015)
24. J. Yang, J. Wright, T. Huang, Y. Ma, Image super-resolution via sparse representation. IEEE Trans. Image Process. **19**, 2861–2873 (2010)
25. R. Timofte, V. Smet, L.V. Gool, A+: adjusted anchored neighborhood regression for fast super-resolution. Lect. Notes Comput. Sci. **9006**, 111–126 (2014)
26. C. Dong, C. Loy, K. He, X. Tang, Image super-resolution using deep convolutional networks. IEEE Trans. Pattern Anal. Mach. Intell. **38**, 295–307 (2015)
27. W. Freeman, T. Jones, E. Pasztor, Example-based super-resolution. IEEE Comput. Graphics Appl. **22**, 56–65 (2002)
28. R. Timofte, V. De, L.V. Gool, *Anchored Neighborhood Regression for Fast Example-Based Super-Resolution*, in *Proceedings of the IEEE International Conference on Computer Vision* (Sydney, Australia, 2013), pp. 1920–1927
29. G. Freedman, R. Fattal, Image and video upscaling from local self-examples. ACM Trans. Graphics **30**, 474–484 (2011)
30. C. Ledig, L. Theis, F. Huszár, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, W. Shi, *Photo-realistic single image super-resolution using a generative adversarial network*, in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2017), pp. 105–114. https://doi.org/10.1109/CVPR.2017.19
31. K. Zhang, W. Zuo, Y. Chen, D. Meng, L. Zhang, Beyond a gaussian denoiser: residual learning of deep CNN for image denoising. IEEE Trans. Image Process. **26**(7), 3142–3155 (2017). https://doi.org/10.1109/TIP.2017.2662206
32. Y. Zhang, Y. Tian, Y. Kong, B. Zhong, Y. Fu, *Residual Dense Network for Image Super-Resolution*, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2018)
33. C. Yan, Z. Li, Y. Zhang, Y. Liu, X. Ji, Y. Zhang, Depth image denoising using nuclear norm and learning graph model. ACM Trans. Multimedia Comput. Commun. Appl. (2020). https://doi.org/10.1145/3404374
34. C. Yan, T. Teng, Y. Liu, Y. Zhang, H. Wang, X. Ji, Precise no-reference image quality evaluation based on distortion identification. ACM Trans. Multimedia Comput. Commun. Appl. (2021). https://doi.org/10.1145/3468872
35. C. Yan, L. Meng, L. Li, J. Zhang, Z. Wang, J. Yin, J. Zhang, Y. Sun, B. Zheng, Age-invariant face recognition by multi-feature fusion and decomposition with self-attention. ACM Trans. Multimedia Comput. Commun. Appl. (2022). https://doi.org/10.1145/3472810
36. J. Yang, J. Wright, T.S. Huang, Y. Ma, Image super-resolution via sparse representation. IEEE Trans. Image Process. **19**(11), 2861–2873 (2010). https://doi.org/10.1109/TIP.2010.2050625
37. R.Y. Tsai, T.S. Huang, *Multiframe image restoration and registration*, in *Proc. Advances in Computer Vision and Image Processing* (1984), pp. 317–339
38. H. Stark, P. Oskoui, High-resolution image recovery from image-plane arrays, using convex projections. J. Opt. Soc. Am. A Opt. Image Sci. **6**, 1715–1726 (1989)
39. A. Patti, M. Sezan, A. Tekalp, Superresolution video reconstruction with arbitrary sampling lattices and nonzero aperture time. IEEE Trans. Image Process. **6**(8), 1064–1076 (1997). https://doi.org/10.1109/83.605404
40. P.E. Eren, M. Sezan, A. Tekalp, Robust, object-based high-resolution image reconstruction from low-resolution video. IEEE Trans. Image Process. **6**(10), 1446-1451 (1997). https://hdl.handle.net/11511/100397
41. A. Tekalp, M. Ozkan, M. Sezan, *High-resolution image reconstruction from lower-resolution image sequences and space-varying image restoration*, in *Proc. ICASSP-92: 1992 IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 3 (1992), pp. 169–172. https://doi.org/10.1109/ICASSP.1992.226249
42. S. Kim, N. Bose, H. Valenzuela, Recursive reconstruction of high resolution image from noisy undersampled multi-frames. IEEE Trans. Acoust. Speech Signal Process. **38**(6), 1013–1027 (1990). https://doi.org/10.1109/29.56062
43. B.C. Tom, N.P. Galatsanos, A.K. Katsaggelos, *Super-Resolution Imaging* (Springer US, Boston, MA, 2000), chap. Reconstruction of a High Resolution Image from Multiple Low Resolution Images, pp. 73–105. https://doi.org/10.1007/0-306-47004-7_4
44. R. Schultz, R. Stevenson, A Bayesian approach to image expansion for improved definition. IEEE Trans. Image Process. **3**(3), 233–242 (1994). https://doi.org/10.1109/83.287017
45. R. Schultz, R. Stevenson, Extraction of high-resolution frames from video sequences. IEEE Trans. Image Process. **5**(6), 996–1011 (1996). https://doi.org/10.1109/83.503915
46. R. Schultz, R. Stevenson, *Improved definition video frame enhancement*, in *Proc. 1995 International Conference on Acoustics, Speech, and Signal Processing*, vol. 4 (1995), pp. 2169–2172. https://doi.org/10.1109/ICASSP.1995.479905
47. R. Hardie, K. Barnard, E. Armstrong, Joint MAP registration and high-resolution image estimation using a sequence of undersampled images. IEEE Trans. Image Process. **6**(12), 1621–1633 (1997). https://doi.org/10.1109/83.650116
48. L. Kondi, D. Scribner, J. Schuler, A comparison of digital image resolution enhancement techniques. Proc. SPIE AeroSense Conf. **4719**, 220–229 (2002)
49. H. He, L.P. Kondi, Resolution enhancement of video sequences with simultaneous estimation of the regularization parameters, in Proc. Image and Video Communications and Processing, Society of Photo-Optical Instrumentation

Engineers (SPIE) Conference Series, vol. 5022, ed. by B. Vasudev, T.R. Hsing, A.G. Tescher. T. Ebrahimi **2003**, 1123–1133 (2003). https://doi.org/10.1117/12.476613
50. H. He, L. Kondi, *MAP based resolution enhancement of video sequences using a Huber-Markov random field image prior model*, in *Proc. 2003 International Conference on Image Processing*, vol. 2 (2003), pp. II–933. https://doi.org/10.1109/ICIP.2003.1246836
51. E.S. Lee, M.G. Kang, Regularized adaptive high-resolution image reconstruction considering inaccurate subpixel registration. IEEE Trans. Image Process. **12**(7), 826–837 (2003). https://doi.org/10.1109/TIP.2003.811488
52. N.P. Galatsanos, V.Z. Mesarović, R. Molina, A.K. Katsaggelos, Hierarchical Bayesian image restoration from partially known blurs. IEEE Trans. Image Process. **9**(10), 1784–1797 (2000). https://doi.org/10.1109/83.869189
53. A. Zomet, A. Rav-Acha, S. Peleg, *Robust super-resolution*, in *Proc. of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001*, vol. 1 (2001), pp. 645–650. https://doi.org/10.1109/CVPR.2001.990535
54. S. Farsiu, D. Robinson, M. Elad, P. Milanfar, Robust shift and add approach to super-resolution. Proceedings of SPIE — The International Society for Optical Engineering **5203** (2003). https://doi.org/10.1117/12.507194
55. S. Farsiu, M. Robinson, M. Elad, P. Milanfar, Fast and robust multiframe super resolution. IEEE Trans. Image Process. **13**(10), 1327–1344 (2004). https://doi.org/10.1109/TIP.2004.834669
56. L. Rudin, S. Osher, E. Fatemi, Nonlinear total variation based noise removal algorithms. Physica D **60**(1), 259–268 (1992). https://doi.org/10.1016/0167-2789(92)90242-F
57. B. Coll, J.M. Morel, A review of image denoising algorithms, with a new one. SIAM J. Multiscale Modeling Simul. (2005). https://doi.org/10.1137/040616024
58. G. Chantas, S. Nikolopoulos, I. Kompatsiaris, Heavy-tailed self-similarity modeling for single image super resolution. IEEE Trans. Image Process. **30**, 838–852 (2020). https://doi.org/10.1109/TIP.2020.3038521
59. K. Dabov, A. Foi, V. Katkovnik, K. Egiazarian, *Image restoration by sparse 3D transform-domain collaborative filtering*, in *Image Processing: Algorithms and Systems VI*, vol. 6812. International Society for Optics and Photonics (SPIE, 2008), p. 681207. https://doi.org/10.1117/12.766355
60. S. Gu, L. Zhang, W. Zuo, X. Feng, *Weighted Nuclear Norm Minimization with Application to Image Denoising* (2014), pp. 2862–2869. https://doi.org/10.1109/CVPR.2014.366
61. U. Kamilov, C. Bouman, G. Buzzard, B. Wohlberg, Plug-and-play methods for integrating physical and learned models in computational imaging: Theory, algorithms, and applications. IEEE Signal Process. Mag. **40**(1), 85–97 (2023). https://doi.org/10.1109/msp.2022.3199595
62. S. Boyd, N. Parikh, E. Chu, B. Peleato, J. Eckstein, Distributed optimization and statistical learning via the alternating direction method of multipliers. Found. Trends Mach. Learn. **3**, 1–122 (2011). https://doi.org/10.1561/2200000016
63. T. Goldstein, S. Osher, The split Bregman method for l1-regularized problems. SIAM J. Imaging Sci. **2**, 323–343 (2009). https://doi.org/10.1137/080725891
64. U.S. Kamilov, H. Mansour, H.B. Wohlberg, A plug-and-play priors approach for solving nonlinear imaging inverse problems. IEEE Signal Process. Lett. **24**(12), 1872–1876 (2017)
65. N. Yurii, Gradient methods for minimizing composite functions. Math. Program. **140**, 125–161 (2012)
66. C. Yan, Y. Hao, L. Li, J. Yin, A. Liu, Z. Mao, Z. Chen, X. Gao, Task-adaptive attention for image captioning. IEEE Trans. Circuits Syst. Video Technol. **32**(1), 43–51 (2022). https://doi.org/10.1109/TCSVT.2021.3067449
67. M.C. Zerva, L.P. Kondi, Video super-resolution using plug-and-play priors. IEEE Access **12**, 11963–11971 (2024). https://doi.org/10.1109/ACCESS.2024.3355195
68. S.C. Park, M.K. Park, M.G. Kang, Super-resolution image reconstruction: a technical overview. IEEE Signal Process. Mag. **20**(3), 21–36 (2003). https://doi.org/10.1109/MSP.2003.1203207
69. G.K. Chantas, N.P. Galatsanos, N.A. Woods, Super-resolution based on fast registration and maximum a posteriori reconstruction. IEEE Trans. Image Process. **16**(7), 1821–1830 (2007)
70. T. Balke, F.R. Davis, C. Garcia-Cardona, M. McCann, L. Pfister, B. Wohlberg, Scientific computational imaging code (SCICO). J Open Source Softw. (2022). https://doi.org/10.21105/joss.04722
71. L. Wang, Y. Guo, L. Liu, Z. Lin, X. Deng, W. An. Deep video super-resolution using HR optical flow estimation (2020)
72. M. Haris, G. Shakhnarovich, N. Ukita, *Recurrent Back-Projection Network for Video Super-Resolution*, in *2019 CVPR* (2019), pp. 3892–3901. https://doi.org/10.1109/CVPR.2019.00402
73. M. Haris, G. Shakhnarovich, N. Ukita, Deep back-projection networks for super-resolution. CoRR **abs/1803.02735** (2018). http://arxiv.org/abs/1803.02735. 1803.02735
74. Y. Jo, S.W. Oh, J. Kang, S.J. Kim, *Deep Video Super-Resolution Network Using Dynamic Upsampling Filters Without Explicit Motion Compensation*, in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition* (2018), pp. 3224–3232. https://doi.org/10.1109/CVPR.2018.00340
75. M.S.M. Sajjadi, R. Vemulapalli, M. Brown, *Frame-Recurrent Video Super-Resolution*, in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2018)
76. X. Wang, K. Yu, K.C. Chan, C. Dong, C.C. Loy. BasicSR. https://github.com/xinntao/BasicSR (2020)
77. Y. Li, P. Jin, F. Yang, C. Liu, M.H. Yang, P. Milanfar, *COMISR: Compression-Informed Video Super-Resolution*, in *2021 IEEE/CVF International Conference on Computer Vision (ICCV)* (2021), pp. 2523–2532. https://doi.org/10.1109/ICCV48922.2021.00254
78. C. Yan, Y. Sun, H. Zhong, C. Zhu, Z. Zhu, B. Zheng, X. Zhou, Review of omnimedia content quality evaluation. J. Signal Process. **38**(6), 1111–1143 (2022). https://doi.org/10.16798/j.issn.1003-0530.2022.06.001

Publisher's Note