Image Restoration

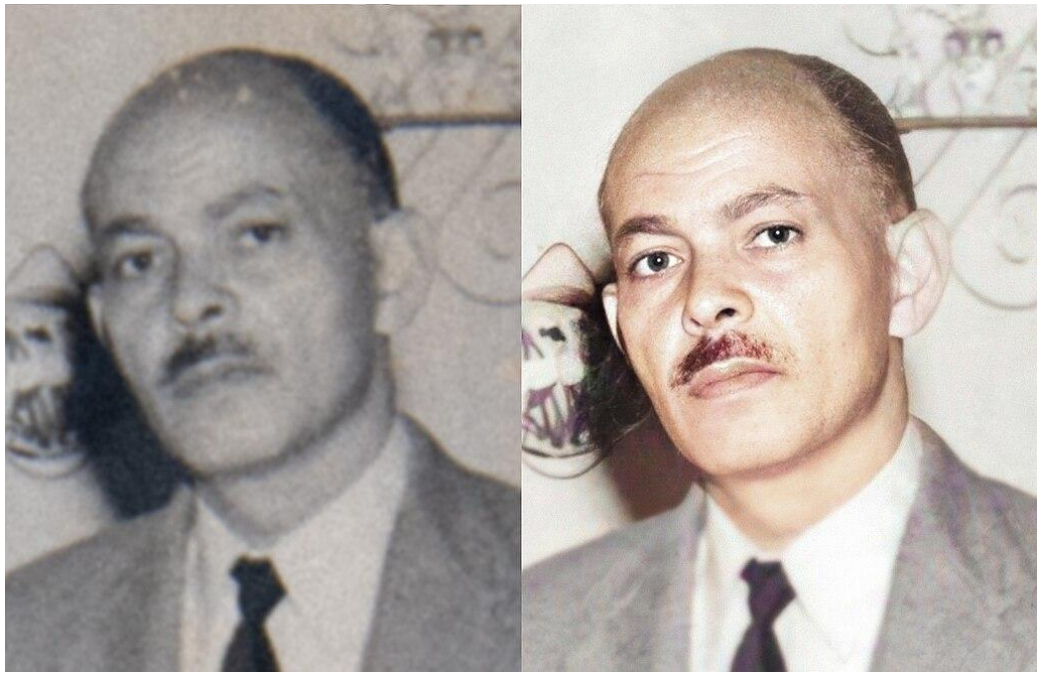


Image restoration and colorization

Image Degradation Model

- Model equation: $y = Hx + n$
  - y: Degraded observation/image.
  - x: Latent (original) image.
  - H: Degradation matrix modeling the process affecting x (e.g., blur).
  - n: Additive noise with zero-mean AWGN and standard deviation v, representing random errors.
- This model forms the basis for tasks like denoising, deblurring, inpainting, and super-resolution.

General Model for Image Restoration Methods

- Objective: Minimize a function combining data fidelity and regularization terms.
- Data fidelity: Measures similarity between restored (x) and observed (y) images.
- Regularization: Imposes constraints based on image priors (e.g., sparsity, low-rankness).

Role of Degradation Matrix H in Image Restoration

- **Denoising**: $H = I$, the identity matrix, indicating no spatial alteration, only the addition of noise n.
- **Deblurring**: H represents convolution with a blur kernel (camera shake, motion blur).
- **Inpainting**: H acts as a mask that removes parts of x, creating gaps in y.
- **Super-resolution**: H models the down-sampling that reduces the resolution of x to produce y.

Introduction to MAP

- Image Restoration (IR) is commonly approached as an ill-posed inverse problem.
- The Maximum A Posteriori (MAP) framework provides a Bayesian perspective for solving such problems.
- The MAP framework effectively combines data consistency with prior knowledge to restore images from degraded observations.

MAP Optimization Problem

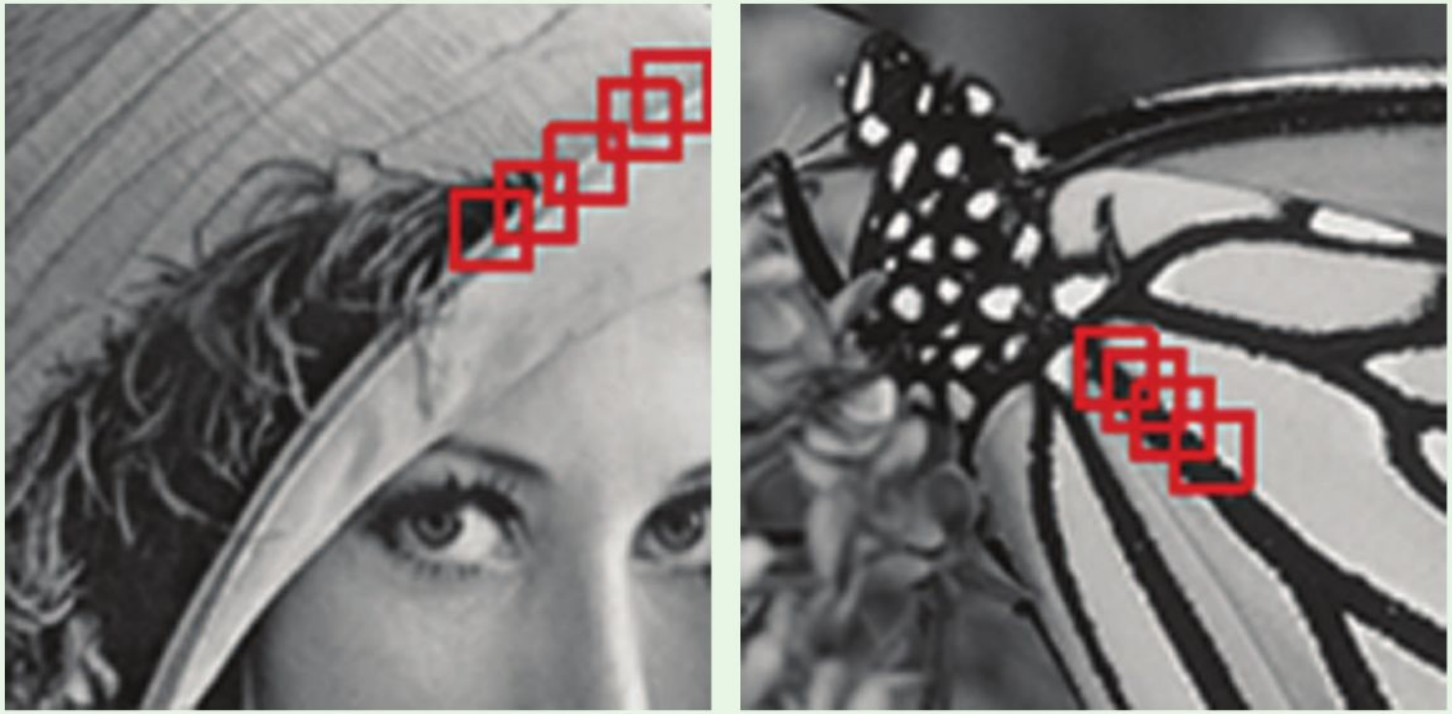
- The MAP estimate $\hat{x}$ is obtained by maximizing the posterior distribution:

$$\hat{x} = \arg\max_{x} \log p(y|x) + \log p(x)$$

- Here, $\log p(y|x)$ represents the log-likelihood of observing y given x, and $\log p(x)$ encodes the prior knowledge of x.

Components of the MAP Framework

- **Fidelity Term**: Represents how close the predicted image Hx is to the observed image y, ensuring consistency with the observed data.
- **Regularization Term:** Encodes prior knowledge about the image, such as smoothness or sparsity, enforcing certain properties on the solution.
- **Trade-off Parameter** λ: Controls the influence of the regularization term relative to the fidelity term.



(a) (b)

The red squares indicate the similar local patterns in an image.

Reformulation to Minimization Problem

- The maximization problem is equivalent to the minimization of the negative log-posterior:

$$\hat{x} = \arg\min_x \frac{1}{2}\|y - Hx\|_2^2 + \lambda\Phi(x)$$

- The solution minimizes an energy function comprising:
  - A fidelity term $\frac{1}{2}\|y - Hx\|_2^2$
  - A regularization term $\lambda\Phi(x)$
- λ is a trade-off parameter balancing the two terms.

Sparse Prior

- **Sparse Representation**: Fundamental principle where images are represented compactly in a transformed domain, exhibiting few significant non-zero coefficients.
- **Transformation Domains**: Emphasizes the use of wavelet, Fourier, or learned dictionaries to find a domain where the image naturally exhibits sparsity.
- **Importance in Denoising**: Sparse prior differentiates between the true image signal (significant coefficients) and noise (sparse and insignificant coefficients), enabling more effective image restoration.

Sparse Representation-based Image Restoration Model

- Equation: $a^* = \arg\min_a \|y - HDa\|_2^2 + \lambda\|a\|_1$
  - a: Sparse coefficients in the transformed domain.
  - D: Overcomplete dictionary for sparse representation.
  - λ: Regularization parameter balancing data fidelity and sparsity.
- Goal: Reconstruct high-quality latent images from degraded observations.

Sparse Prior

- **MAP Framework Integration**: Incorporates the sparse prior assumption into the MAP framework to balance fidelity to observed noisy data with sparsity enforcement.
- **L1 Norm Sparsity Enforcer**: Utilizes the L1 norm of the transformed image coefficients to promote sparsity, minimizing non-zero elements while preserving essential image features.

Low-Rank Minimization Overview

Let Y be a matrix representing degraded image patches. The goal is to recover the latent low-rank matrix X which can be formulated as a nuclear norm minimization problem:

$$\hat{X} = \arg\min_{X} \|Y - X\|_F^2 + \lambda \|X\|_* \tag{1}$$

where:

- $\|\cdot\|_F$ denotes the Frobenius norm.
- $\|\cdot\|_*$ represents the nuclear norm, i.e., the sum of the singular values.
- λ is the regularization parameter.

Low-Rank Prior

- **Low-Rank Concept**: Captures the essence that real-world images can often be represented efficiently with matrices of lower rank, indicating redundancy and strong correlations among pixels.
- For example in Image Denoising, it utilizes the inherent structure and correlations within images to facilitate noise reduction while preserving significant image details.
- **MAP Framework Compatibility**: Low rank-prior is comparable with (MAP) framework by incorporating structural assumptions about images into the denoising process.

Low-Rank Prior

- **Incorporating Low-Rank Prior**: Introduces a regularization term (often the nuclear norm as a **convex surrogate** for rank) to penalize high-rank solutions, promoting simpler, cleaner images.
- **Optimization Challenges**: Addresses the non-convexity and discontinuity of direct rank optimization by leveraging the nuclear norm, enabling efficient, gradient-based optimization methods.
- **Practical Implications**: performs good image restoration by ensuring that the reconstructed image honors both the observed data and the low-rank structure indicative of natural images.

Classical vs DL-Based Image Restoration

Aspect	Classical	DL
Generality to handle different IR problems	General	Limited
Optimization process time	Time-consuming	Efficient
Data-driven end-to-end learning	Absent	Present
Generality of learned models	Present	Limited
Interpretability of learned models	Present	Limited

Denoising Task Learning Objective

- Objective: $f = \arg\min_f \mathbb{E}_y[f(x) - y]_2^2$
  - f: Function approximating the inverse of the corrupting process.
  - $\mathbb{E}_y$: Expectation over the distribution of original images.

Skip Connections and Residual Learning (Network Structure Prior)

- Skip Connections:
  - Enhance the network's ability to recover detailed features by linking convolutional and deconvolutional layers.
- Residual Learning:
  - Models the difference (residual) between corrupted and clean images, focusing on learning the "corruption" for improved restoration.

SRGAN

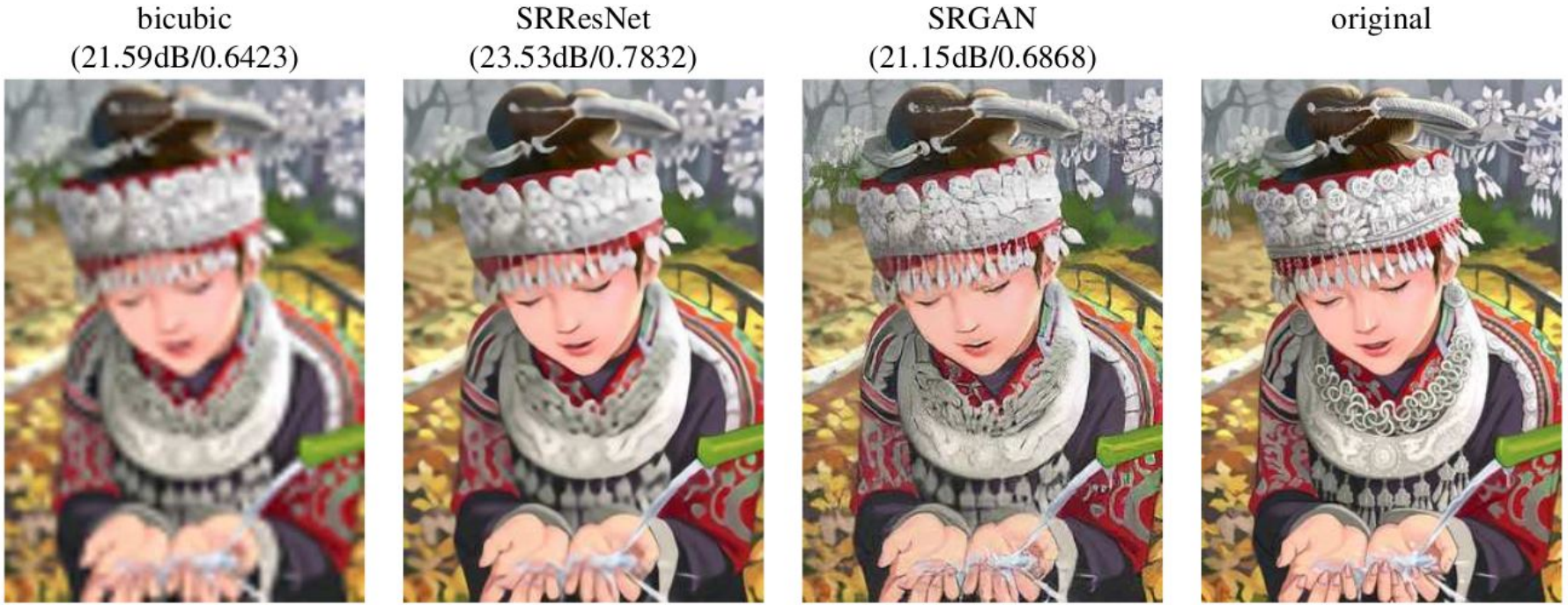


Figure 2: From left to right: bicubic interpolation, deep residual network optimized for MSE, deep residual generative adversarial network optimized for a loss more sensitive to human perception, original HR image. Corresponding PSNR and SSIM are shown in brackets. [4× upscaling]

SRGAN for Image Super-Resolution

- Introduces SRGAN for achieving photo-realistic results in 4× image upscaling.
- Utilizes a perceptual loss function that combines adversarial and content loss.
- Overcomes limitations of MSE-based optimization by focusing on perceptual quality.

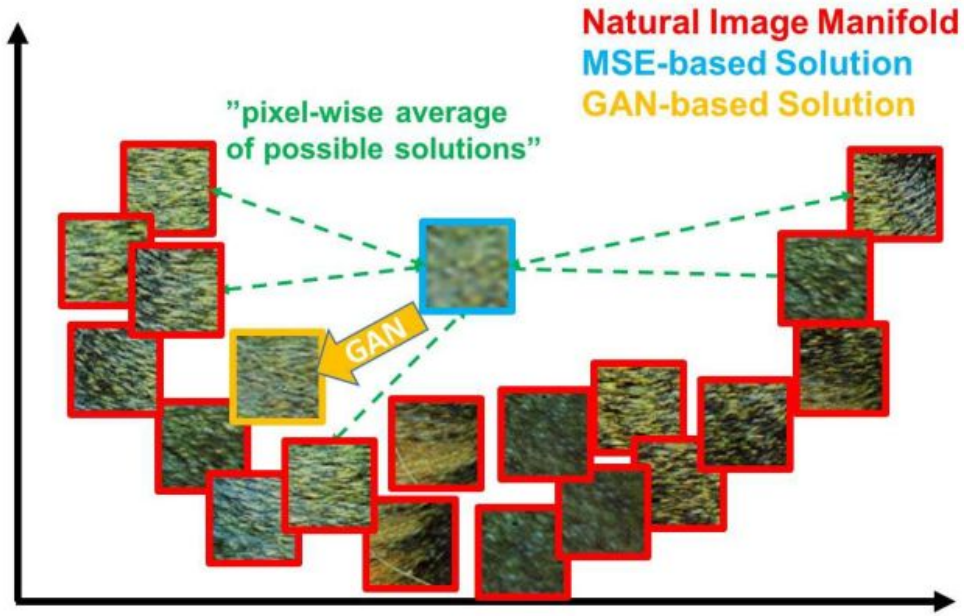


Figure 3: Illustration of patches from the natural image manifold (red) and super-resolved patches obtained with MSE (blue) and GAN (orange). The MSE-based solution appears overly smooth due to the pixel-wise average of possible solutions in the pixel space, while GAN drives the reconstruction towards the natural image manifold producing perceptually more convincing solutions.

SR Optimization Problem

$$\hat{\theta}_G = \arg\min_{\theta_G} \frac{1}{N} \sum_{n=1}^{N} l^{SR}(G_{\theta_G}(I_n^{LR}), I_n^{HR}) \tag{3}$$

- Aims to find generator parameters that minimize the SR-specific loss.
- Balances between generating high-resolution images and staying true to target high-resolution images.

Adversarial Training

$$\min_{\theta_G} \max_{\theta_D} \mathbb{E}_{I^{HR} \sim p_{\text{train}}(I^{HR})}[\log D_{\theta_D}(I^{HR})] + \mathbb{E}_{I^{LR} \sim p_G(I^{LR})}[\log(1 - D_{\theta_D}(G_{\theta_G}(I^{LR})))] \quad (4)$$

- Engages generator and discriminator in a minimax game.
- Enhances the generation of indistinguishable high-resolution images.

Perceptual Loss Function

$$l^{SR} = l_X^{SR} + 10^{-3} l_{Gen}^{SR} \tag{5}$$

- Combines content loss (for perceptual similarity) with adversarial loss.
- Aims for photo-realism in super-resolved images.

Content Loss Variants

MSE Content Loss:

$$l^{SR}_{MSE} = \frac{1}{r^2WH}\sum_{x=1}^{rW}\sum_{y=1}^{rH}(I^{HR}_{xy} - G_{\theta_G}(I^{LR})_{xy})^2 \tag{6}$$

VGG Loss:

$$l^{SR}_{VGG/i.j} = \frac{1}{W_{ij}H_{ij}}\sum_{x=1}^{W_{ij}}\sum_{y=1}^{H_{ij}}(\phi_{ij}(I^{HR})_{xy} - \phi_{ij}(G_{\theta_G}(I^{LR}))_{xy})^2 \tag{7}$$

- Leverages deep features and perceptual similarity for content loss.

Adversarial Loss

$$l_{Gen}^{SR} = -\sum_{n=1}^{N} \log D_{\theta_D}(G_{\theta_G}(I^{LR})) \tag{8}$$

- Measures the generator's success in fooling the discriminator.

Here, $D_{\theta_D}(G_{\theta_G}(I^{LR}))$ is the probability that the reconstructed image $G_{\theta_G}(I^{LR})$ is a natural HR image. For better gradient behavior we minimize $-\log D_{\theta_D}(G_{\theta_G}(I^{LR}))$ instead of $\log[1 - D_{\theta_D}(G_{\theta_G}(I^{LR}))]$.

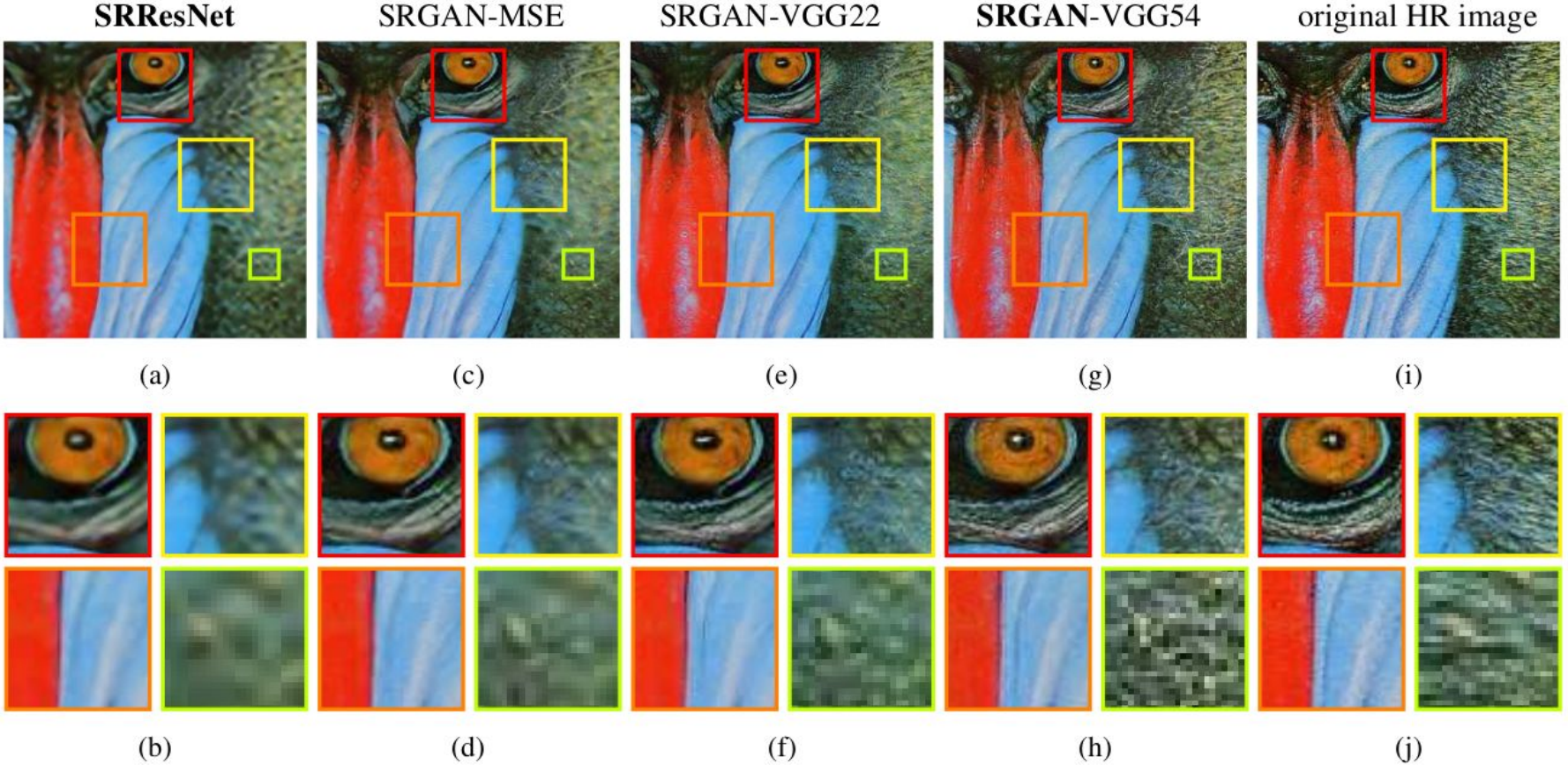


Figure 6: **SRResNet** (left: a,b), SRGAN-MSE (middle left: c,d), SRGAN-VGG2.2 (middle: e,f) and **SRGAN**-VGG54 (middle right: g,h) reconstruction results and corresponding reference HR image (right: i,j). [4× upscaling]

Conclusion

- SRGAN sets new benchmarks for photo-realistic super-resolution.
- Employs perceptual and adversarial loss to enhance visual quality.
- Challenges conventional metrics like PSNR, emphasizing perceptual quality.
- Future directions include exploring invariant content loss functions and efficient architectures.

DeblurGAN

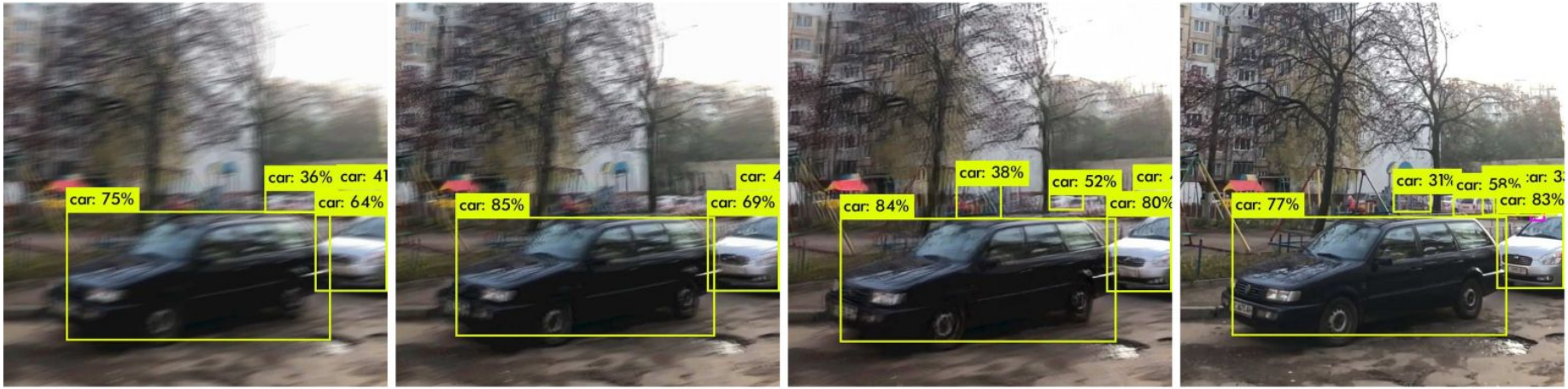
DeblurGAN for Motion Deblurring

- Introduces DeblurGAN as an end-to-end solution for motion deblurring.
- Achieves superior structural similarity and visual appearance.
- Provides a significant speed advantage over competitors.
- Enhances dataset realism through novel augmentation techniques.

Image Deblurring Model

$$I^B = k(M) * I^S + N \tag{9}$$

- I^B: Blurred image.
- $k(M)$: Blur kernels dependent on motion field M.
- I^S: Desired sharp image.
- N: Additive noise.



(a) Blurred photo (b) Nah *et al.* [25] (c) DeblurGAN (d) Sharp photo

Figure 9: YOLO object detection before and after deblurring

DeblurGAN Overview

- Utilizes conditional GANs for blind motion deblurring.
- Employs a multi-component loss including perceptual loss.
- Innovates in dataset generation for training.
- Enhances evaluation with object detection-based protocols.

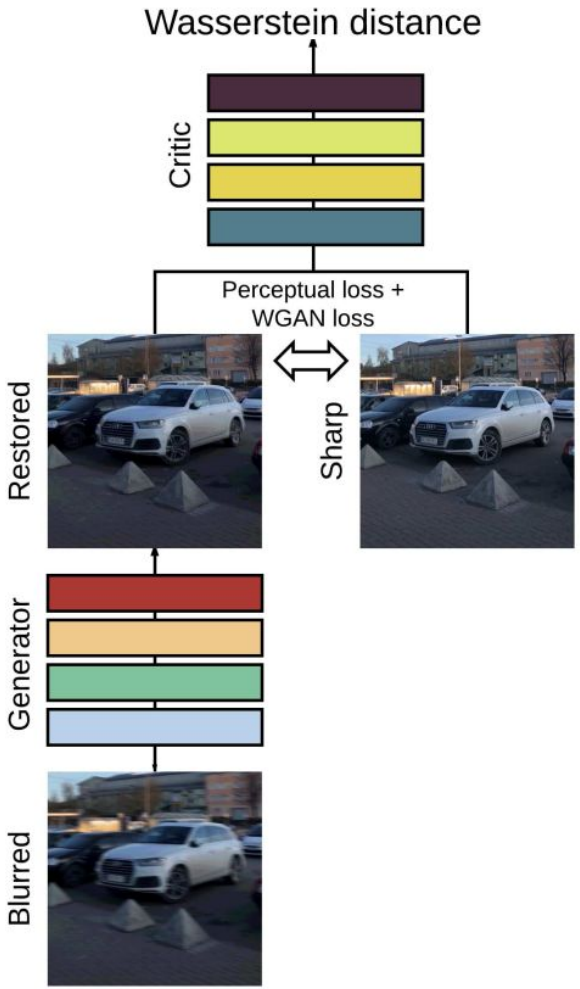


Figure 4: DeblurGAN training. The generator network takes the blurred image as input and produces the estimate of the sharp image. The critic network takes the restored and sharp images and outputs a distance between them. The total loss consists of the WGAN loss from critic and the perceptual loss [17]. The perceptual loss is the difference between the VGG-19 [34] *conv3.3* feature maps of the sharp and restored images. At test time, only the generator is kept.

GAN Objective

$$\min_G \max_D \mathbb{E}_{x \sim P_r}[\log(D(x))] + \mathbb{E}_{\tilde{x} \sim P_g}[\log(1 - D(\tilde{x}))] \tag{10}$$

- Balances generation and discrimination for realistic deblurring.

WGAN Objective with Gradient Penalty (WGAN-GP)

- Optimizes GAN training stability via Earth-Mover distance.
- Reduces sensitivity to generator architecture.
- Promotes smooth and continuous gradient updates.

Wasserstein GAN Critic

- In WGAN, Discriminator function *D* is called critic, that converts an image into a score.
- The critic no longer emits a simple probability. The critic is used to approximate the Earth-Mover Distance (Wasserstein Distance) between the fake and real distributions.

WGAN Objective Function

$$\min_{G} \max_{D \in \mathcal{D}} \mathop{\mathbb{E}}_{\boldsymbol{x} \sim \mathbb{P}_r} \left[D(\boldsymbol{x}) \right] - \mathop{\mathbb{E}}_{\tilde{\boldsymbol{x}} \sim \mathbb{P}_g} \left[D(\tilde{\boldsymbol{x}}) \right]$$

- The equation shows that objective function computes the difference between the expectations from real samples and generated samples.
- Here D is the set of **1-Lipschitz functions**.

Gulrajani, Ishaan, et al. "Improved training of wasserstein gans." Advances in neural information processing systems (2017).

1-Lipschitz Function

We say that this function is 1-Lipschitz if it satisfies the following inequality for any two input images, x_1 and x_2:

$$\frac{|\, D(x_1) - D(x_2) \,|}{|\, x_1 - x_2 \,|} \leq 1$$

Here, $x_1 - x_2$ is the average pixel-wise absolute difference between two images and

$|\, D(x_1) - D(x_2) \,|$ is the absolute difference between the critic predictions.

The Lipschitz Constraint force the discriminator output to not take on extreme values (exploding gradients).

WGAN Key Idea

- In the WGAN paper, the authors show how it is possible to enforce the **Lipschitz constraint by clipping the weights** of the critic to lie within a small range, [–0.01, 0.01], after each training batch.
- The WGAN is trained using labels of 1 for real and –1 for fake.

Adversarial and Perceptual Loss

$$L_{GAN} = -\sum_{n=1}^{N} D_{\theta_D}(G_{\theta_G}(I^B)) \tag{12}$$

$$L_{\text{perceptual}} = \frac{1}{W_{ij}H_{ij}} \sum_{x=1}^{W_{ij}} \sum_{y=1}^{H_{ij}} (\phi_{ij}(I^S)_{xy} - \phi_{ij}(G_{\theta_G}(I^B))_{xy})^2 \tag{13}$$

- Adversarial loss measures generator's success in fooling discriminator.
- Perceptual loss focuses on high-level image content for realism.

Loss Function Composition

$$L = L_{GAN} + \lambda \cdot L_X \tag{11}$$

- Combines adversarial and content loss for optimal deblurring.
- λ adjusts the balance between loss components.

Figure 2: GoPro images [25] processed by DeblurGAN. Blurred – left, DeblurGAN – center, ground truth sharp – right.

Key Takeaways

- DeblurGAN introduces an effective approach for motion deblurring using cGANs.
- Combines WGAN-GP with perceptual loss for enhanced texture restoration.
- Demonstrates advancement in speed, dataset realism, and evaluation metrics.
- Sets new benchmarks in motion deblurring performance and application.

Context Encoders

Context Encoders: Feature Learning by Inpainting

- Context Encoders (CE) are a form of convolutional neural networks trained to predict missing parts of a scene from their surroundings.
- The method aims to learn visual features unsupervisedly by generating the contents of an arbitrary image region based on its context.
- CE uses an encoder-decoder architecture where the encoder captures the context into a latent feature representation, and the decoder generates the missing content.
- Training incorporates both pixel-wise reconstruction loss and an adversarial loss to ensure generated content blends seamlessly with the original image.

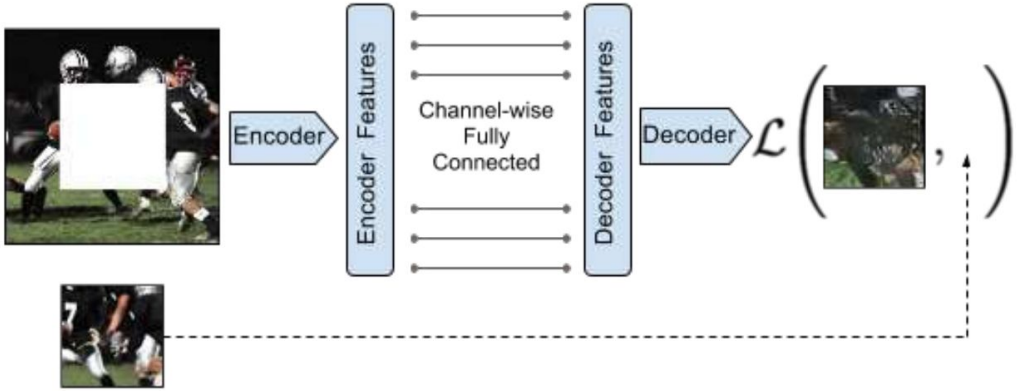
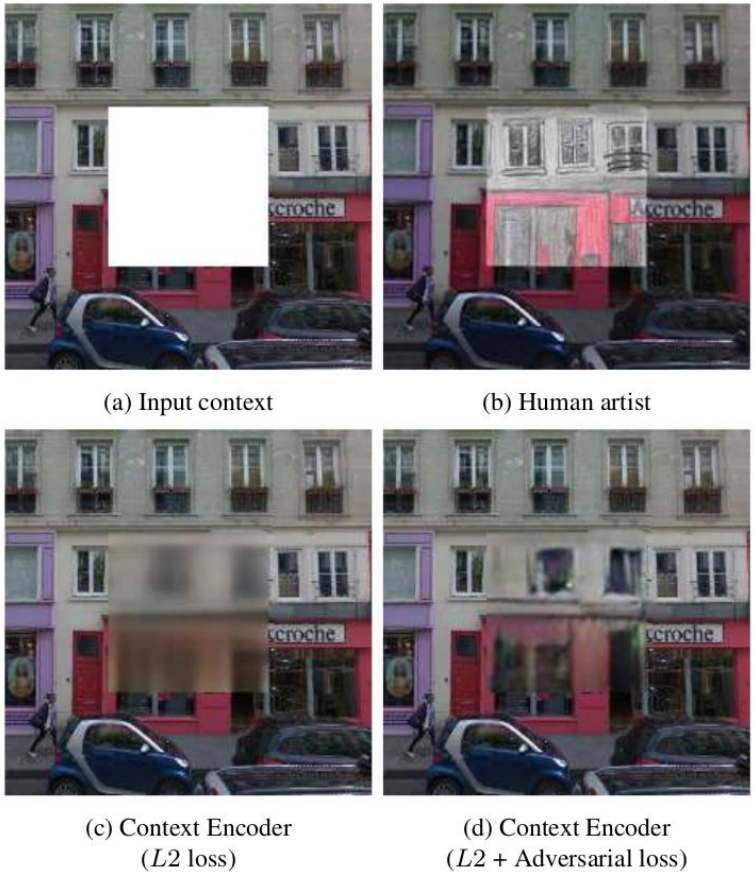


Figure 2: Context Encoder. The context image is passed through the encoder to obtain features which are connected to the decoder using channel-wise fully-connected layer as described in Section 3.1. The decoder then produces the missing regions in the image.



(a) Input context

(b) Human artist

(c) Context Encoder ($L2$ loss)

(d) Context Encoder ($L2$ + Adversarial loss)

Key Components

- **Encoder-Decoder Architecture:** Facilitates the understanding of the image context and generation of the missing part.

- **Reconstruction and Adversarial Loss:** Ensures the generated content is not only structurally coherent but also visually plausible.

Mathematical Formulation

- **Reconstruction Loss:**

$$L_{\text{rec}} = \|M \odot (x - F((1 - M) \odot x))\|_2^2$$

- **Adversarial Loss:**

$$L_{\text{adv}} = \max_D \mathbb{E}_{x \sim p_{\text{data}}(x)}[\log D(x)] + \mathbb{E}_{\tilde{x} \sim p_{\tilde{x}}(\tilde{x})}[\log(1 - D(F((1 - M) \odot x)))]$$

- **Joint Loss:**

$$L = \lambda_{\text{rec}} L_{\text{rec}} + \lambda_{\text{adv}} L_{\text{adv}}$$

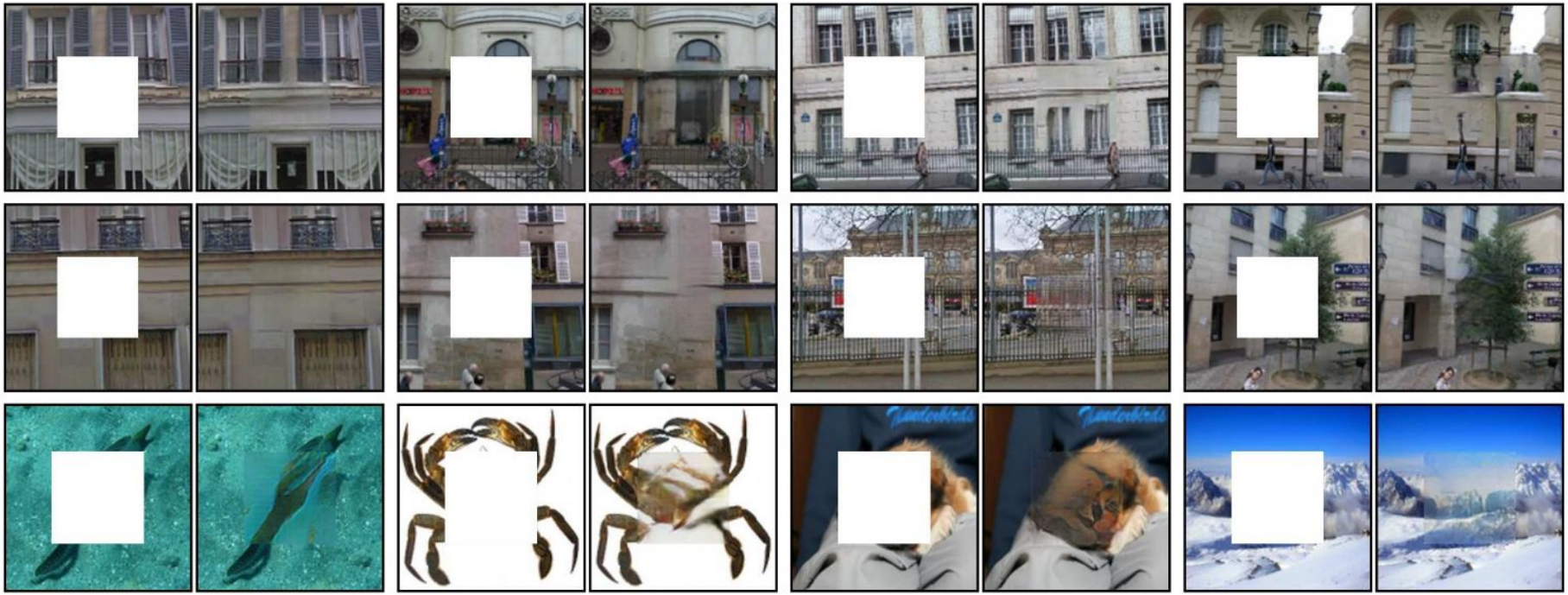


Figure 4: Semantic Inpainting results for context encoder trained jointly using reconstruction and adversarial loss.

Conclusion

- They are capable of filling in missing image regions in a contextually and visually coherent manner.
- Training the model in an unsupervised manner allows for significant flexibility and application across various domains without the need for labeled data.
- Context Encoders not only improve inpainting results but also contribute to the field of unsupervised feature learning, demonstrating the utility of generative models in understanding image content.