Single-Image Super-Resolution for Satellite Imagery

Thomas Chabal[1,2] & Quentin Spinat[1,2]

[1]Ecole des Ponts ParisTech
Applied Mathematics and Computer Science Department

[2]E.N.S. Paris-Saclay
Master MVA

Deep Learning Project, January 2021

Introduction

- Need for better resolution in many fields
  - Biology
  - Satellite Imagery
  - Photography
- Yet, physical improvements of cameras are more and more difficult
- Drives us towards the development of other methods

Problem Definition

- **Given** an image I^{LR} of **low-resolution**
- **Obtain a higher resolution image $\widehat{I}^{HR}$ upscaled by a factor** s

$$U : I^{LR} \in \mathbb{R}^{w} \times \mathbb{R}^{h} \rightarrow \widehat{I}^{HR} \in \mathbb{R}^{w \times s} \times \mathbb{R}^{h \times s}$$

Problem Definition

- **Given** an image I^{LR} of **low-resolution**
- **Obtain a higher resolution image $\widehat{I}^{HR}$ upscaled by a factor** s

$$U : I^{LR} \in \mathbb{R}^w \times \mathbb{R}^h \to \widehat{I}^{HR} \in \mathbb{R}^{w \times s} \times \mathbb{R}^{h \times s}$$

- Fine reconstruction goal : for a function U,

$$\forall I^{LR}, \left\| \widehat{I}^{HR} - I^{HR} \right\| \to 0$$

Methodology

Classical Computer Vision : spline interpolations

- Yet, poor quality (due to Fourier domain)

With the rise of neural networks, new solutions :

- **Autoencoder** : compress the input image into a latent representation and then decode it
- **(Robust) U-Net** : use prior information on the input image to refine the resulting upscaled image
- **Generative Adversarial Networks** : generate plausible details in images and make them seem real
- Yet, unexact methods that may hallucinate

Training and evaluation framework: dataset and transformations

- Dataset used : **Massachussetts Road Dataset**
  - Satellite images
  - $\approx$ 800 training images, 13 test images
- Transforms :
  - Dataset originally used for road detection
  - Apply downscales and then upscales
  - For RUNet: additional random gaussian blur between downscaling and upscaling

Training and evaluation framework: metrics

Metrics :

- L2 loss
- Peak Signal-to-Noise Ratio (PSNR)
  - Metric for noise reduction, assesses quality of images
- Structural Similarity Index Metric (SSIM)
  - More global overview of image quality than PSNR
- Perceptual loss
  - VGG-16 features extracted
  - L2 distance in this feature representation space
  - Better metric empirically compared to visual impression

Classical Computer Vision

Spline interpolations :

- Exact methods

$$\forall (i,j) \in [|0; w|] \times [|0; h|], \qquad U(I^{LR})_{s\times i, s\times j} = I^{HR}_{s\times i, s\times j}$$

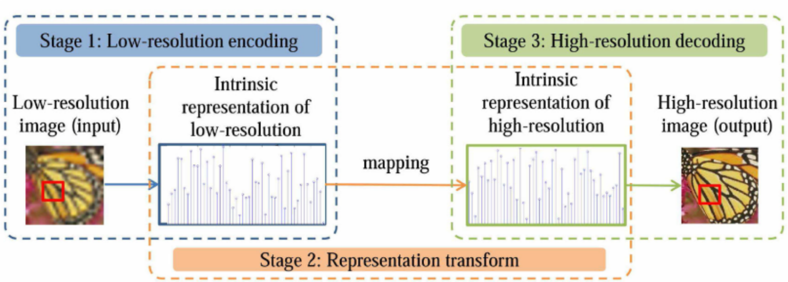
- Yet, slow to compute
- Don't affect Fourier domain size, so no details (=high frequencies) added

$$\varphi_{n-spline} = \bigstar_{i=1}^{n} 1_{[-\frac{1}{2};\frac{1}{2}[^2}$$

where $\bigstar$ is the convolution product.

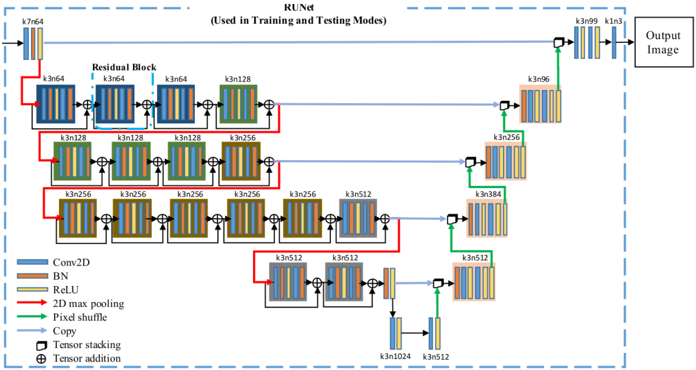
Coupled Deep Autoencoder

- Two Autoencoders trained together
- Idea: get low dimensional representation of high and low resolution images, then map them together
- We reimplemented it from scratch



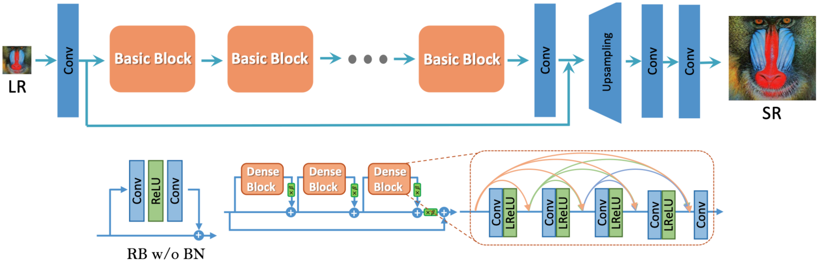
Robust UNet

- U-Net architecture
- Architectural differences : Residual blocks in encoding part
- Training specification : Blur input images before upscaling them
- We reimplemented it from scratch



ESRGAN

- Generate details in images
- Output = mix between GAN output and PSNR-oriented network output



Evaluation: metrics

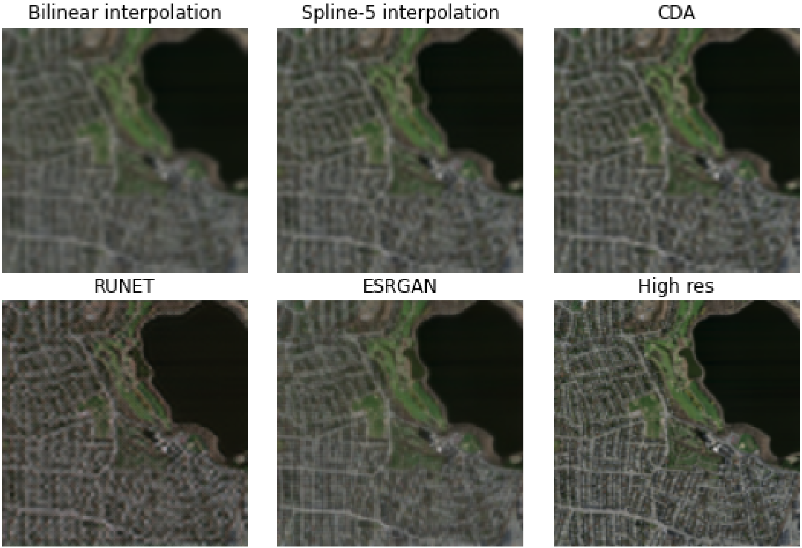
	MSE (10^{-3})	SSIM (10^{-1})	PSNR	Perceptual
Bilinear	3.025	8.621	22.56	6.388
Spline-5	2.925	9.126	24.46	5.547
CDA	2.522	**9.281**	**25.17**	5.152
RUNET	3.486	8.903	22.52	5.931
GAN	**2.033**	9.125	25.03	**4.858**

Table 1: Quantitative results over the test set

Evaluation: visual results

Evaluation: visual results

Evaluation: visual results



Conclusion

- Classical interpolations of poor quality, yet exact
- Great visual results with RUNet, ESRGAN could be finetuned on our dataset to give better outputs
- Metrics not always related to image quality $\rightarrow$ Find a better metric
- Further exploration for higher upscaling factors
- Depending on the application of satellite imagery, need to have exact images and no hallucination, so prefer classical vision