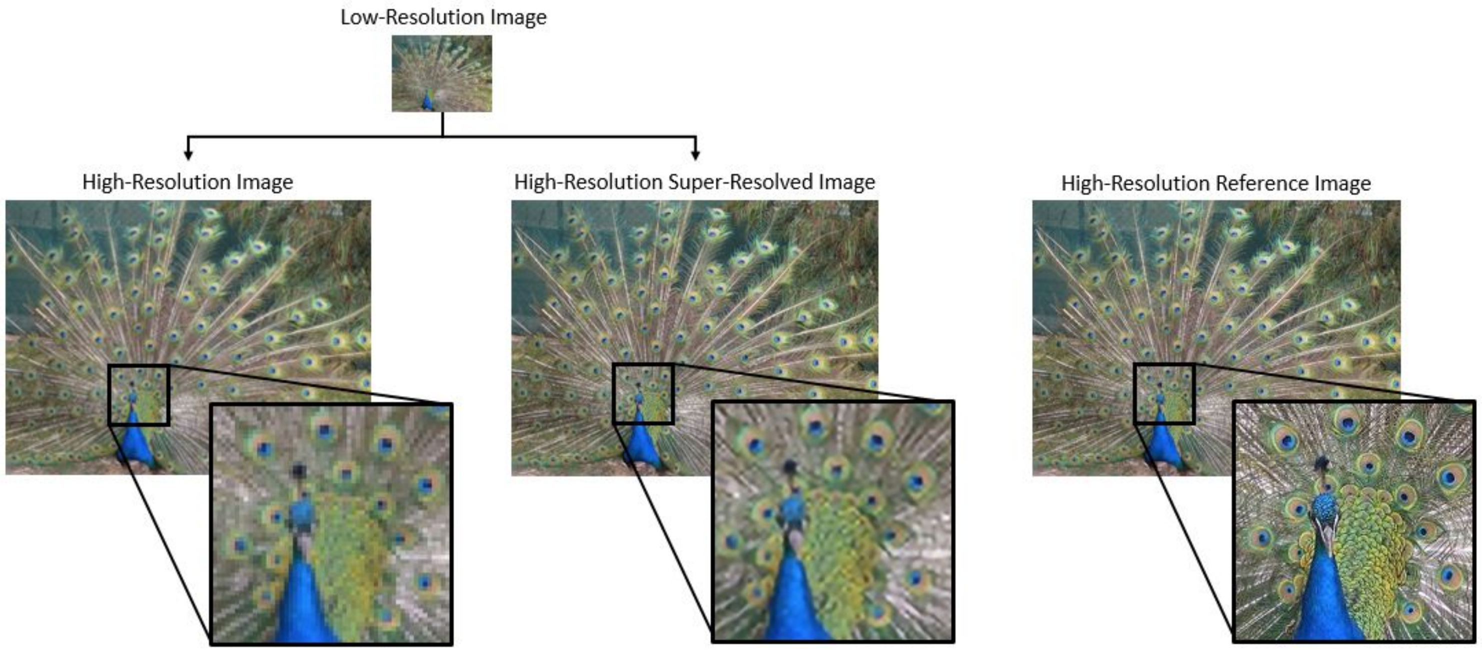
Single Image Super-resolution from Transformed Self-Exemplars



Team #39:

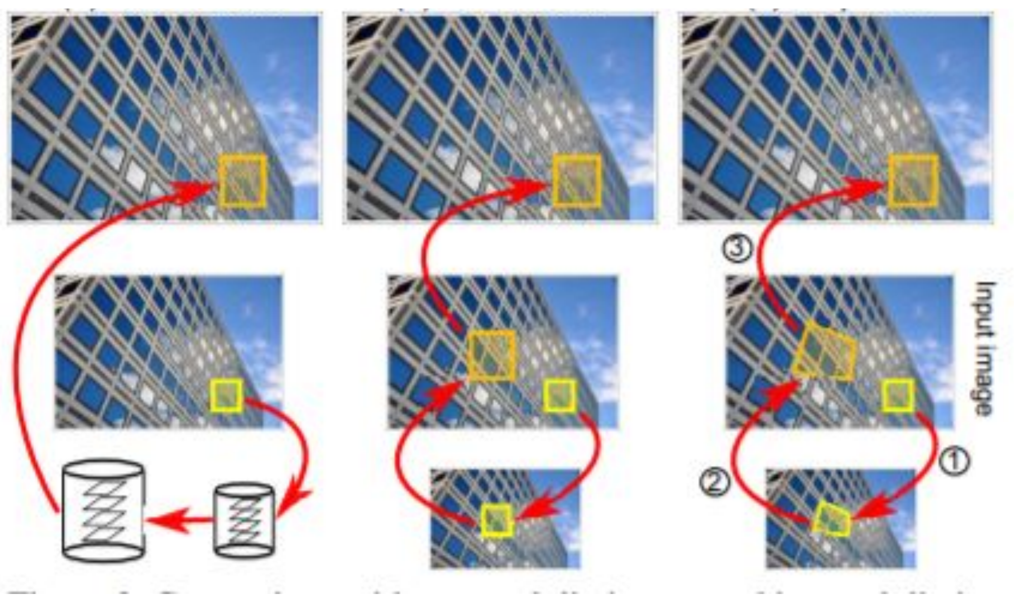
Shaforostov Artem
Lazareva Elizaveta
Aleshina Anna
Myshak Anton
Serkova Ekaterina

Problem statement:

- Most modern single image super-resolution (SR) methods rely on machine learning techniques.
- These methods focus on learning the relationship between low-resolution (LR) and high-resolution (HR) image patches. A popular class of such algorithms uses an external database of natural images as a source of LR-HR training patch pairs. Existing methods have employed various learning algorithms for learning this LR to HR mapping, including nearest neighbor approaches , manifold learning, dictionary learning, locally linear regression, and convolutional networks.
- However, methods that learn LR-HR mapping from external databases have certain shortcomings. The number and type of training images required for satisfactory levels of performance are not clear. Large scale training sets are often required to learn a sufficiently expressive LR-HR.

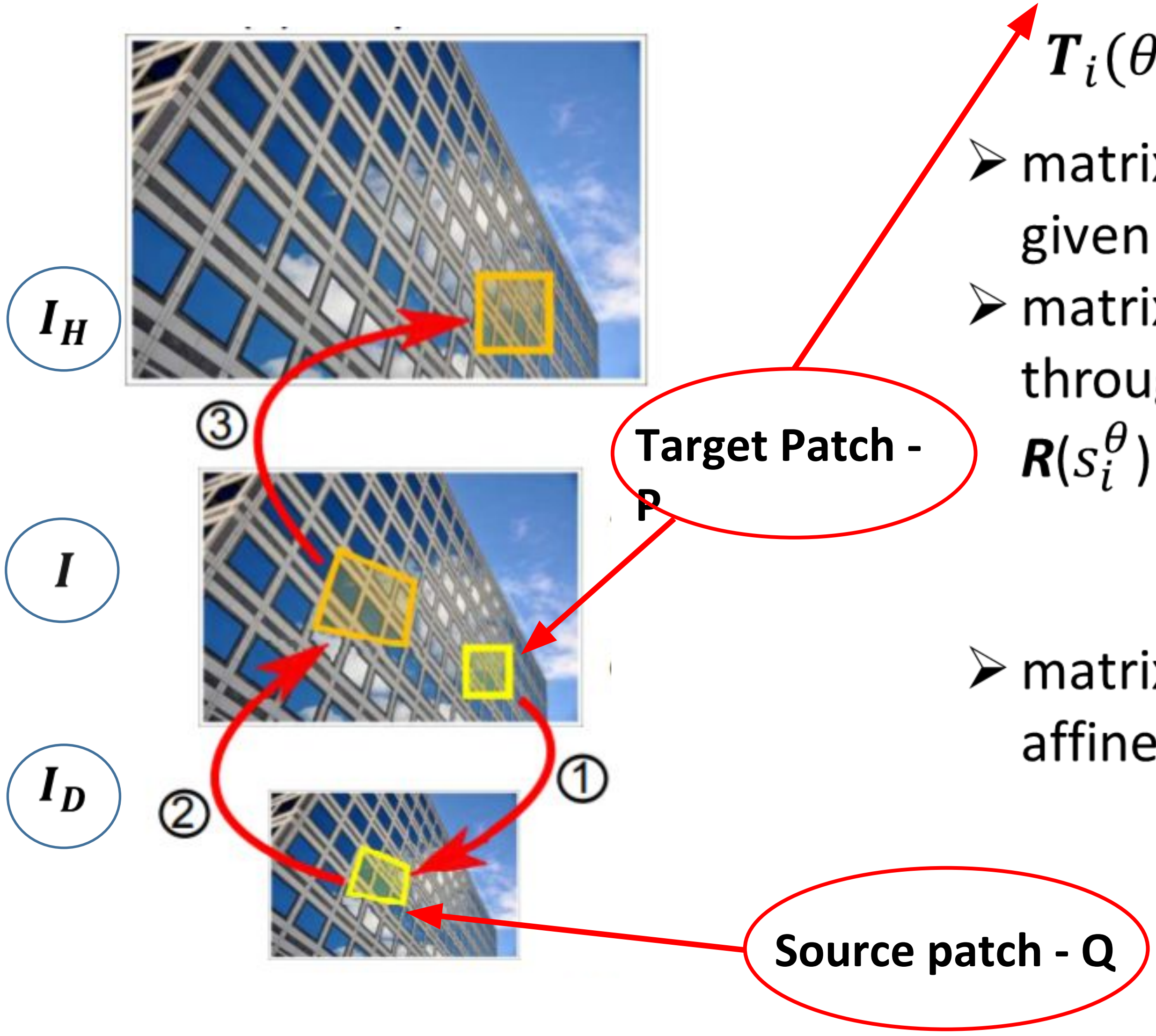
We propose a self-similarity driven SR algorithm that expands the internal patch search space.

Input low resolution image *I* → Subsample + Blur → Get downsample → PatchMatch, Nearest Neighbor field approaches → HR-sample



Super-resolution scheme:

Transformation matrix T: warps P to its best matching patch Q (source patch) in I_D



$$T_i(\theta_i) = \boldsymbol{H}\left(t_i, s_i^x, s_i^y, m_i\right)\boldsymbol{S}\left(s_i^s, s_i^\theta\right)\boldsymbol{A}\left(s_i^\alpha, s_i^\beta\right)$$

- matrix $\boldsymbol{H}$ — captures the perspective deformation given P and Q positions and the planer parameter
- matrix $\boldsymbol{S}$ - captures the similarity transformation through a scaling parameter s_i^s and a rotation matrix $\boldsymbol{R}(s_i^\theta)$:

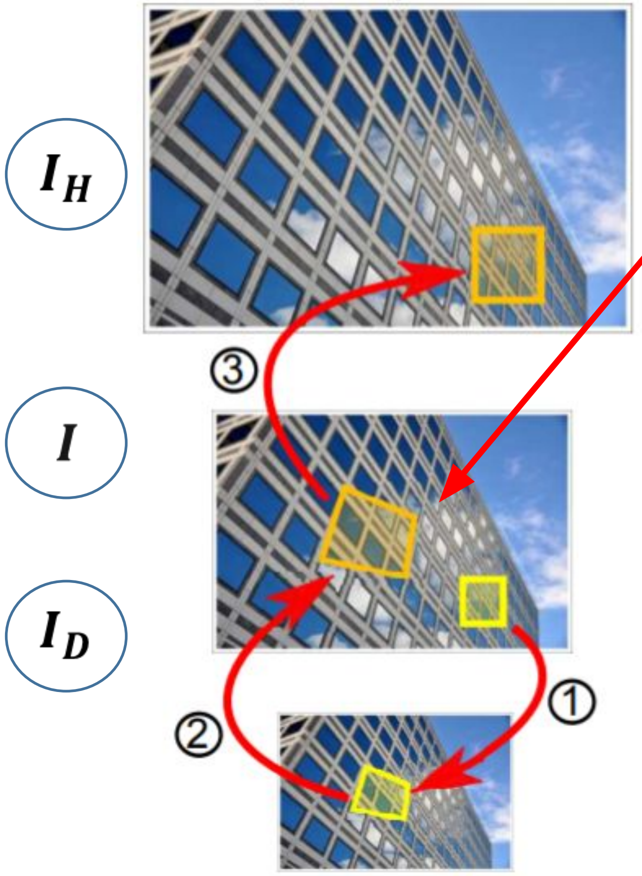
$$\boldsymbol{S}\left(s_i^s, s_i^\theta\right) = \begin{bmatrix} s_i^s \boldsymbol{R}(s_i^\theta) & 0 \\ 0 & 1 \end{bmatrix}$$

- matrix $\boldsymbol{A}$ - captures the shearing mapping in the affine transformation:

$$\mathbf{A}\left(s_i^\alpha, s_i^\beta\right) = \begin{bmatrix} 1 & s_i^\alpha & 0 \\ s_i^\beta & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Super-resolution scheme:

To obtain the parameters of such a transformation, we estimate a nearest neighbor field (NNF) between *I* and *I_D* using a PatchMatch algorithm with some modifications:



Ω - Set of pixel indices of the input LP image *I*.

For each target patch $P(\boldsymbol{t}_i)$ centered at position $\boldsymbol{t}_i = (t_i^x, t_i^y)^T$ in *I*, our goal is to estimate a transformation matrix T_i that maps the target patch $P(t_i)$ to its nearest neighbor in the downsampled image I_D.

We have a field of transformations parametrized by θ_i for each i^{th} pixel in the input LR image.

Our objective NNF function includes 3 costs:

$$\min_{\{\theta_i\}} \sum_{i \epsilon \Omega} \mathbf{E}_{app}(t_i, \theta_i) + \boldsymbol{E}_{plane}(t_i, \theta_i) + \boldsymbol{E}_{scale}(t_i, \theta_i)$$

Nearest Neighbors field:

where SRF indicates the desired SR factor, e.g., 2x, 3x, or 4x, and the function Scale(·) indicates the scale estimation of a projective transformation matrix.

$$\min_{\{\theta_i\}} \sum_{i\epsilon\Omega} \mathbf{E}_{app}(t_i,\theta_i) + \boldsymbol{E}_{plane}(t_i,\theta_i) + \boldsymbol{E}_{scale}(t_i,\theta_i)$$

Appearance cost:

Similarity between the sampled target and source patches. Our metric is Gaussian-weighted sum-of squared distance in the RGB space:

$$\mathbf{E}_{app}(t_i,\theta_i) = \|W_i(P(t_i) - Q(t_i,\theta_i)\|_2^2$$

Scale cost:

matching target patches to itself in the downsampled image I_D

$$\mathbf{E}_{scale} = \lambda_{sclae}\min(0, SRF - Scale(\boldsymbol{T}_i),$$

Plane compatibility cost:

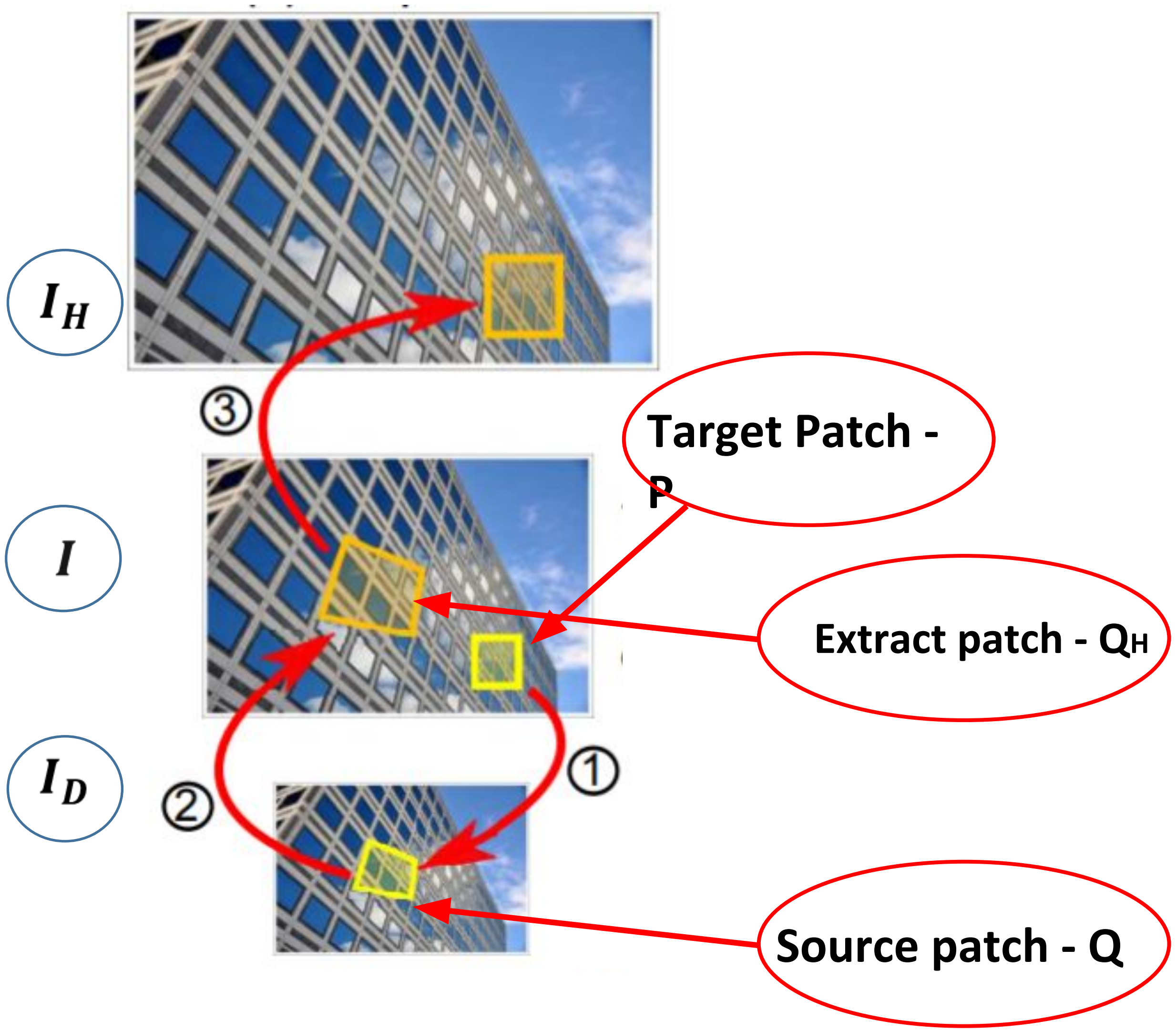
For images, we can often localize planes in the scene using standard vanishing point detection techniques.

$$\mathbf{E}_{plane} = -\lambda_{plane}\log(Pr[m_i|(s_i^x,s_i^y)] \times Pr[m_i|(t_i^x,t_i^y)]),$$

$Pr[m_i|(x,y)]$ is the posterior probability of assigning lab

at pixel position (x,y)

Super-resolution scheme:



- ✓ After we extract Q_H from the image I, which is the HR version of the source patch Q.
- ✓ We use the inverse of the computed transformation matrix T to 'unwarp' the HR patch Q_H, to obtain the selfexemplar P_H, which is our estimated HR version of the target patch P.
- ✓ We paste P_H in the HR image I_H at the location corresponding to the LR patch P.
- ✓ We repeat the above steps for all target patches to obtain an estimate of the HR image I_H.

Obtained results:

Low Resolution input images

Vanishing points of images

High Resolution input images

Obtained results:

Obtained results:

Metric	**Scale**	**Bicubic**	**Ours**
PSNR(set 5)	4x	22.31	24.34
SSIM(set 5)	4x	0.6634	0.7123

Conclusions:

- Our method effectively increases the size of the limited internal dictionary by allowing geometric transformation of patches. We achieve state-of-the-art results without using any external training images.
- We propose a decomposition of the geometric patch transformation model into:
  - (i) perspective distortion for handling structured scenes
  - (ii) additional affine transformation for modeling local shape deformation.
- We use and make available a new dataset of urban images containing structured scenes as a benchmark for SR evaluation.