Sparsity vs Data-Driven — Image Super-Resolution Methods

Abijith Jagannath Kamath, Manogna S
Department of Electrical Engineering
Indian Institute of Science, Bangalore
Email: {abijithj, manognas}@iisc.ac.in

This presentation is submitted in partial fulfilment for the requirements of E9 241 Digital Image Processing, August 2020, course project.

Super-Resolution and Upsampling

Conventional Methods for Upsampling Sequences

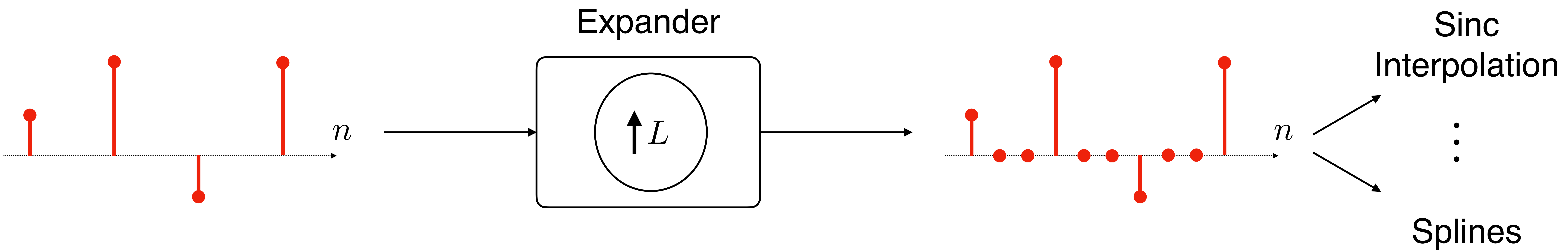


Figure 1: Upsampling discrete sequence using an expander followed by interpolation.

Image Super-Resolution Methods

- Classical interpolation based methods
- Super-resolution as an inverse problem
- Data-driven methods

Sparsity Priors on Images

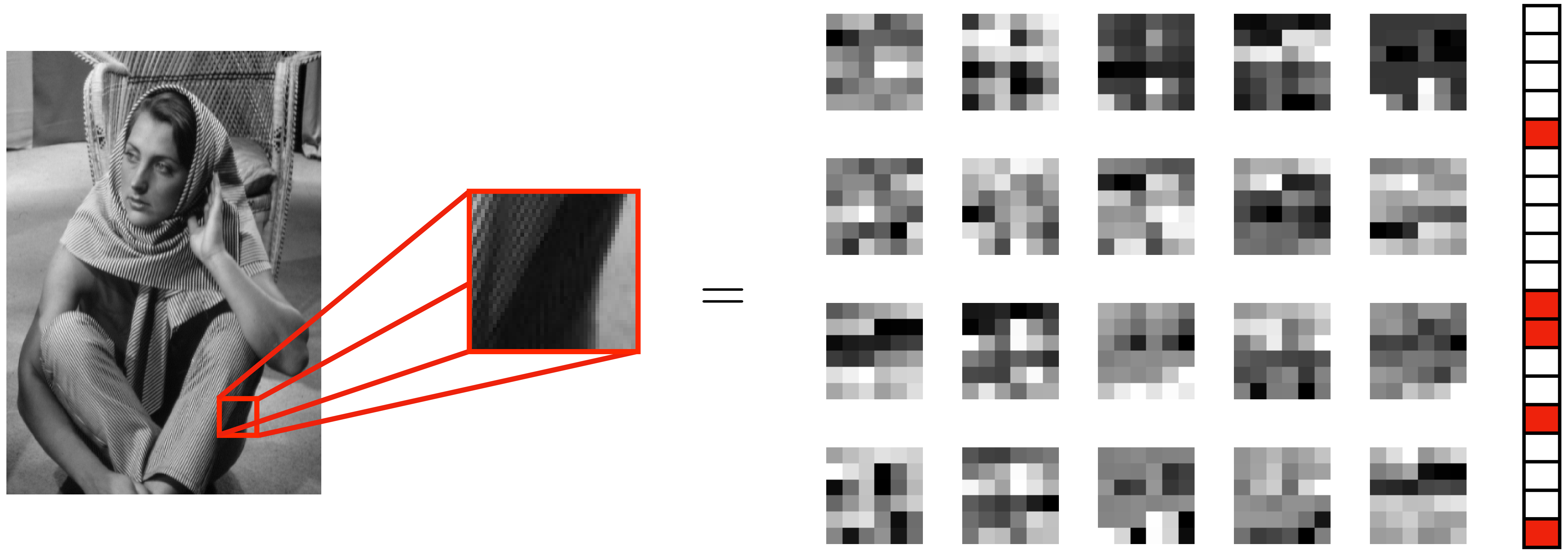


Figure 2: Sparse representation of image patches.

- A vector **y** is compressible if

$$\exists \boldsymbol{D} \in \mathbb{R}^{n \times k},\ \mathbf{y} = \boldsymbol{D}\boldsymbol{\alpha},$$

  where $k \gg n$ and $\|\boldsymbol{\alpha}\|_0 \ll k$.

- Images are known to be compressible.

- Sparse representation is assumed to be invariant to scale,

$$\text{if } \mathbf{y} = \boldsymbol{D}_\ell \boldsymbol{\alpha},\ \exists \boldsymbol{D}_h \in \mathbb{R}^{N \times k}, \mathbf{x} = \boldsymbol{D}_h \boldsymbol{\alpha},$$

  where $k \gg N > n$. **x** is a higher resolution version of **y**.

Super-Resolution using Sparsity

Formulation

- Let **X** be the high-resolution image. The low resolution image:

$$\mathbf{Y} = SH\mathbf{X} \qquad \text{(Reconstruction Constraint)}$$

  where H is a blur kernel and S is a downsampling operator.

- Patches **x** of **X** are compressible on some overcomplete dictionary $\boldsymbol{D}_h \in \mathbb{R}^{N\times k}$:

$$\mathbf{x} = \boldsymbol{D}_h\boldsymbol{\alpha}, \text{for some } \|\boldsymbol{\alpha}\|_0 \ll k \qquad \text{(Sparsity Constraint)}$$

Super-Resolution using Sparsity

Proposed Method [1]

- The sparse representation can be obtained from **Y** by solving:

$$\underset{\boldsymbol{\alpha}}{\text{minimise}} \; \|\boldsymbol{D}_\ell \boldsymbol{\alpha} - \mathbf{y}\|_2^2 + \lambda \|\boldsymbol{\alpha}\|_1, \qquad \text{(Sparse Step)}$$

  where **y** are patches of the **Y** and $\boldsymbol{D}_\ell$ is the dictionary corresponding to the low-resolution image.

- The solution is projected to satisfy reconstruction constraint by solving:

$$\underset{\mathbf{X}}{\text{minimise}} \; \|SH\mathbf{X} - Y\|_2^2 + c\|\mathbf{X} - \mathbf{X}_0\|_2^2, \qquad \text{(Regularisation)}$$

  where $\mathbf{X}_0$ is the solution obtained from the sparse step.

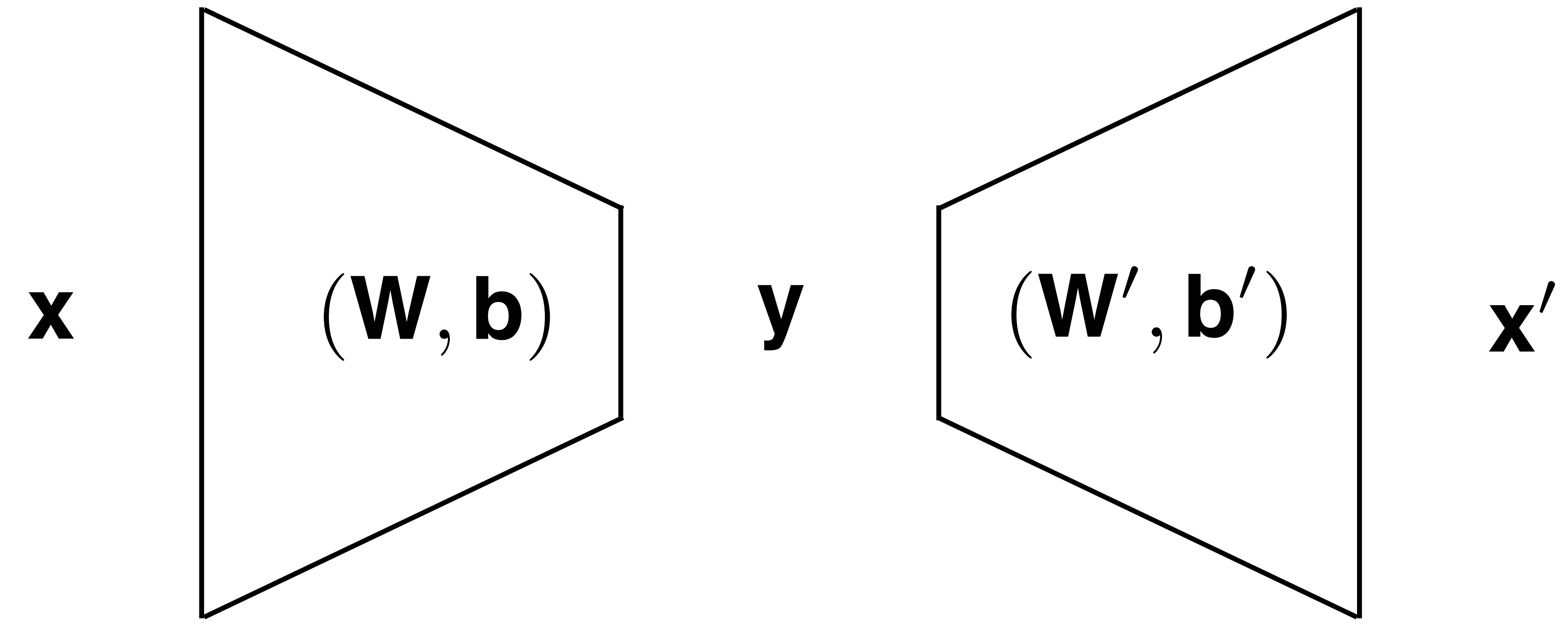
Dictionary Learning

- Dictionaries $\boldsymbol{D}_\ell$ and $\boldsymbol{D}_h$ are learned using training.
- MOD or K-SVD maybe used at the dictionary learning step.

[1] J. Yang et al., "Image Super-Resolution via Sparse Representation," *Trans. Image Process.*, 2010.

Super-Resolution using Convolutional Autoencoders

- Images can be represented as a sparse vector in a very high-dimensional space.
- Equivalently, images can be represented in a dense low-dimensional space.
- The mapping $\mathbf{y} = f(\mathbf{W} * \mathbf{x} + \mathbf{b})$ is learnt using a convolutional neural network.



- An image can be reconstructed by considering the decoder network.
- The decoder map $\mathbf{x}' = f'(\mathbf{W}' * \mathbf{y} + \mathbf{b}')$ is a deconvolutional network.
- The output of the deconvolutional network has larger dimensions, thus achieving image super-resolution.

Network Architecture

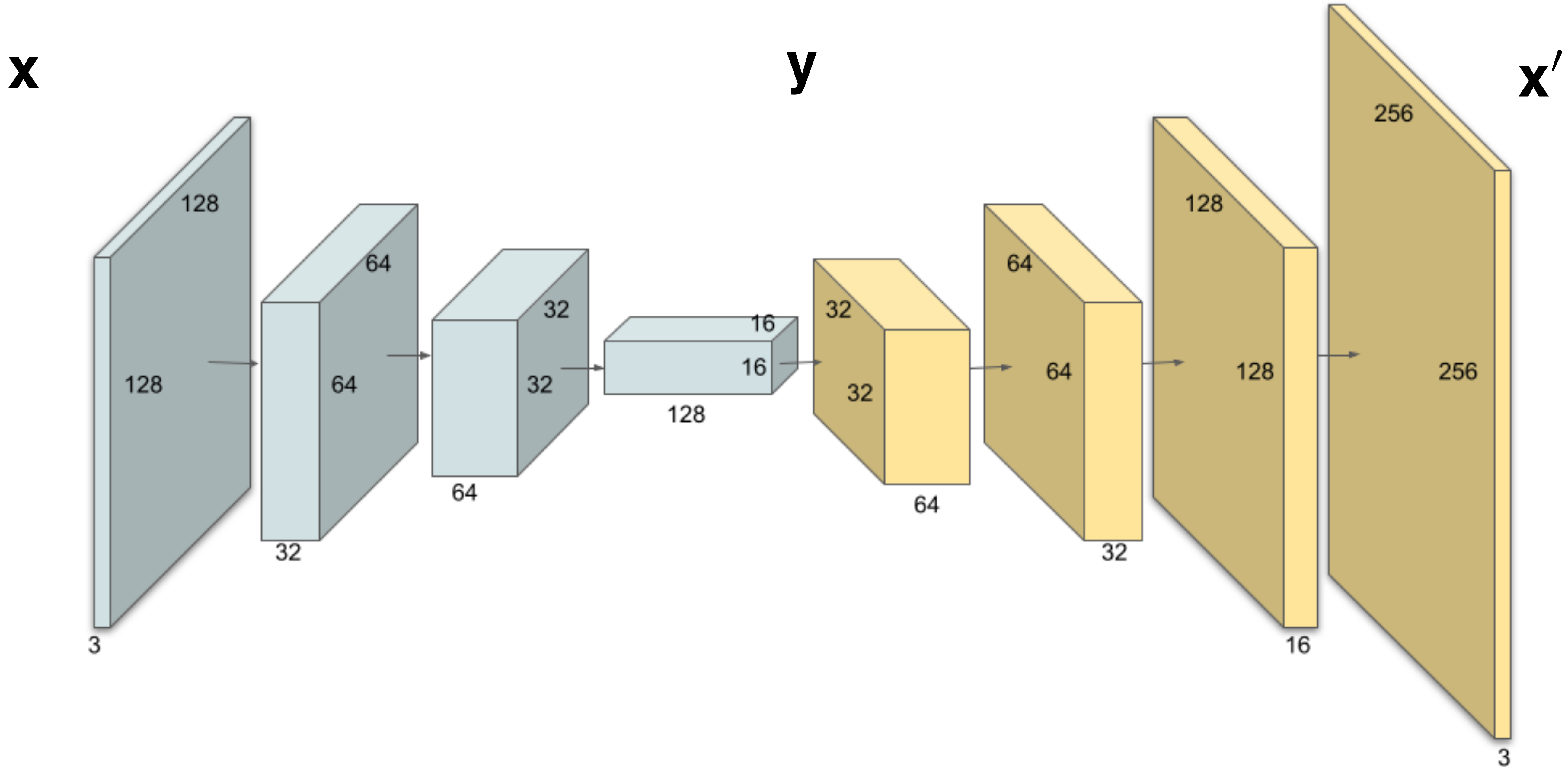


Figure 3: Autoencoder network for 128×128 RGB image.

- Encoder network: 3 convolution layers, stride 2, filter size 3×3 with channels $[32, 64, 128]$.
- Decoder network: 4 deconvolution layers, stride 2, filter size 3×3 with channels $[64, 32, 16, 3]$.

[2] P. Vincent et al., “Extracting and Composing Robust Features with Denosing Autoencoders,” *ICML.*, 2008.

Network Training

- ~ 15000 image patches in the training set [3].
- 32×32 size low-resolution and 64×64 size high-resolution patch pairs used for training.
- Mean-squared error loss function:

$$L(\theta) = \frac{1}{n} \sum_{i=1}^{N} \|\mathbf{x} - \mathbf{x}'\|^2,$$

  where $\theta = \{\mathbf{W}, \mathbf{W}', \mathbf{b}, \mathbf{b}'\}$ are the learning parameters.
- Adam optimiser for 300 epochs with constant learning rate 0.001.

[3] C. Dong et al., "Image Super-Resolution Using Deep Convolutional Networks," *Trans. Pattern Anal. Mach. Intell.*, 2015.

Results: Super-Resolution

Super-Resolution by 2

Bird

ScSR: 36.80dB

SdA-3: 32.59dB

Monarch

ScSR: 27.43dB

SdA-3: 31.07dB

Face

ScSR: 34.14dB

SdA-3: 30.50dB

Image	*ScSR (PSNR in dB)*	*SdA-3 (PSNR in dB)*
Baby	**36.4635**	34.5258
Bird	**36.8057**	32.5930
Monarch	27.4369	**31.0713**
Face	**34.1473**	30.5024
Woman	**32.0706**	26.6857

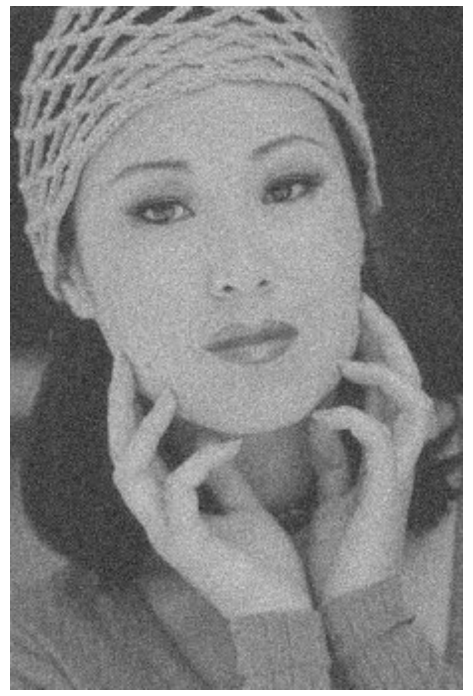
Table 1: Super-resolution using ScSR vs. SdA-3 on Set5.

Results: Noise Analysis

Face
$\sigma = 4$

ScSR

SdA-3



Woman
$\sigma = 25$

ScSR

SdA-3

Noise Std.	4		8		16		25	
Image	*ScSR*	*SdA-3*	*ScSR*	*SdA-3*	*ScSR*	*SdA-3*	*ScSR*	*SdA-3*
Baby	30.2659	**33.8602**	24.3241	**32.6868**	19.7995	**30.5509**	17.6192	**28.6378**
Bird	29.464	**32.1479**	23.3559	**31.0317**	19.7085	**28.5978**	17.2755	**26.3496**
Monarch	**31.3241**	26.3983	**27.0462**	26.2999	22.3257	**25.4620**	19.5528	**23.9455**
Face	30.1463	**30.7285**	23.8293	**30.0731**	19.2218	**28.7201**	15.6224	**27.3790**
Woman	**31.2331**	30.4201	25.622	**29.8606**	21.1032	**28.6236**	17.3557	**26.8349**

Table 2: Comparison denoising performance of ScSR vs. SdA-3 for images in Set5 dataset.

Conclusions

Comparisons

- ScSR outperforms SdA-3 in the noise-free setting.
- SdA-3 is more robust to noise compared to ScSR.
- SdA-3 introduces denoising artefacts.
- ScSR and SdA-3 are scale-invariant.
- SdA-3 readily extends to multichannel inputs.

Future Work

- SdA-3 with deeper network.
- SdA-3 training on a larger dataset.
- Introducing skipped connections in the SdA-3 network.
- Using K-SVD or other faster/robust methods for ScSR dictionary learning.

Thank You!