Optimisation of Convolutional Neural Networks for Super-Resolution on Face Images

June 28, 2016

Andreas Aakerberg, Malte Pedersen, Christoffer B. Rasmussen
16gr843@es.aau.dk

Vision, Graphics and Interactive Systems

Aalborg University

AALBORG UNIVERSITY
DENMARK

Agenda

Introduction

Super Resolution

Problem Statement

Dataset

Deep Learning

Super Resolution Convolutional Neural Network

Research

Conclusion

Future work

Introduction

The number of surveillance cameras in public areas keeps increasing. There are now approx. 600,000 cameras in Denmark [1].

- The quality of the recorded footage is, however, still limited by various factors.
  - External factors such as lighting conditions and occlusion.
  - Internal factors such as compression artefacts, noise, limited resolution and wide FOV.

[1] Source: http://www.dr.dk/nyheder/indland/600000-kameraer-danskerne-er-de-mest-overvaagede

Introduction

Identification of persons can be difficult.

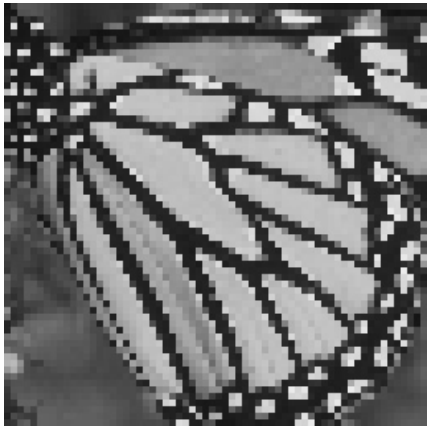
2016 Brussels Airport bombings

- The technical quality of surveillance cameras in Denmark is generally poor[2].

[2] C. Østergaard, “Daarlige kameraer kan svække nyt register”, Ingeniøren feb 2016

Super Resolution

Increase spatial resolution to create a more detailed image.



LR image

HR image

- Direct solution: Increase pixel density of the image sensor.
- Super-Resolution: Transform one or more LR images into a HR image.

Super Resolution

Pros (P) and cons (C) of increasing sensor resolution vs. performing Super Resolution

Increasing sensor resolution

- **P** - Actual increase of resolution/details.
- **C** - Expensive and cumbersome.
- **C** - Higher requirements to storage capacity.
- **C** - Can introduce noise and blur.

Performing SR

- **P** - A generic cost effective solution.
- **P** - Can be used only when needed to limit storage needs.
- **C** - Enhancement is a "guess".
- **C** - SoTA methods are still limited.

Super Resolution

Super-Resolution methods:

- Multiple image methods, based on sub-pixel shifts between LR-images.
- Single image methods, typically learning based.

SoTA Super-Resolution: Deep Learning, scientific computing accelerated using GPUs.

[3]

[3] Image source: http://www.nvidia.com/object/tesla-workstations.html

Problem Statement

Based on the SoTA super-resolution convolutional neural network, SRCNN, by Dong et al., we aim to investigate the following:

How can a state-of-the-art CNN based SR algorithm be enhanced to create images of higher quality?

- How does a SRCNN optimised to specific image types, such as faces, perform?
- How does a SRCNN combined with a classic SR method perform?

Dataset

Training datasets of face images are needed as the SRCNN is a learning based super-resolution method.

	AR500	**AToF**	**LFD**
Characteristics	Similar conditions	Very large variation	Large variation
Images	500	153	13300

Several test datasets are also used (AR5, TDRF and SET5).

Dataset

AR500

AToF

LFD

Deep Learning
Overview

Convolutional Neural Networks

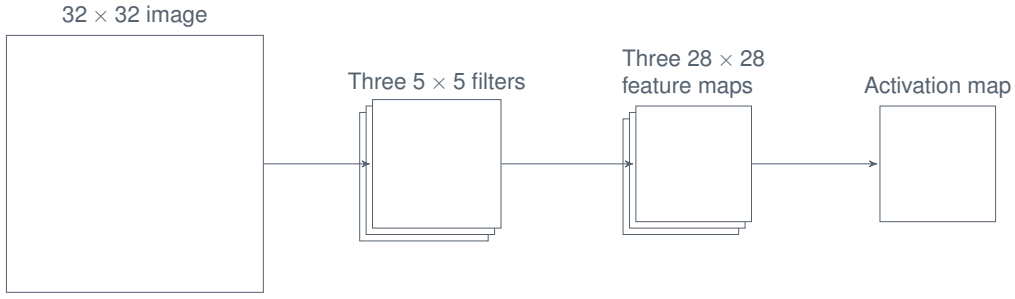
- Based on Artificial Neural Networks
- Originally presented in 1979 [4]
  - Renaissance due to GPU computing

[4] K. Fukushima, "Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by a shift in position," Biological Cybernetics, vol. 36, no. 4, pp. 193-202, 1980

Deep Learning
CNN Architecture

Layers

- Convolutional
- Activation
  - ReLU, Sigmoid
- Final
  - Classification: Fully-connected output scores
  - Regression: Single output activation map



CNN architecture

Deep Learning
CNN Architecture

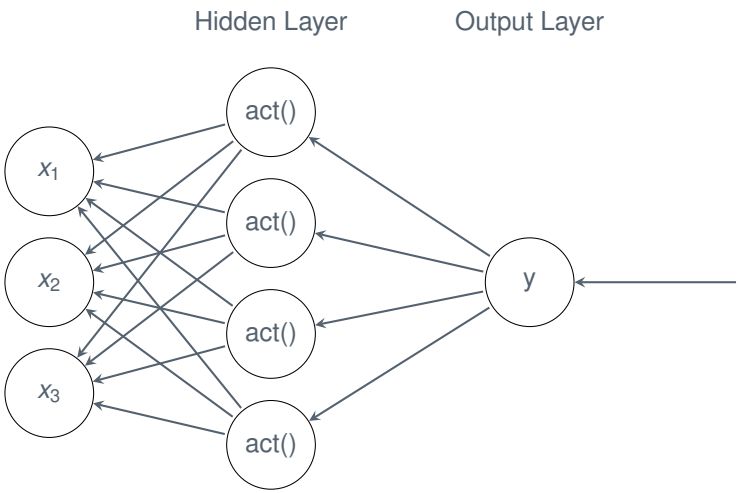
Convolutional Layer

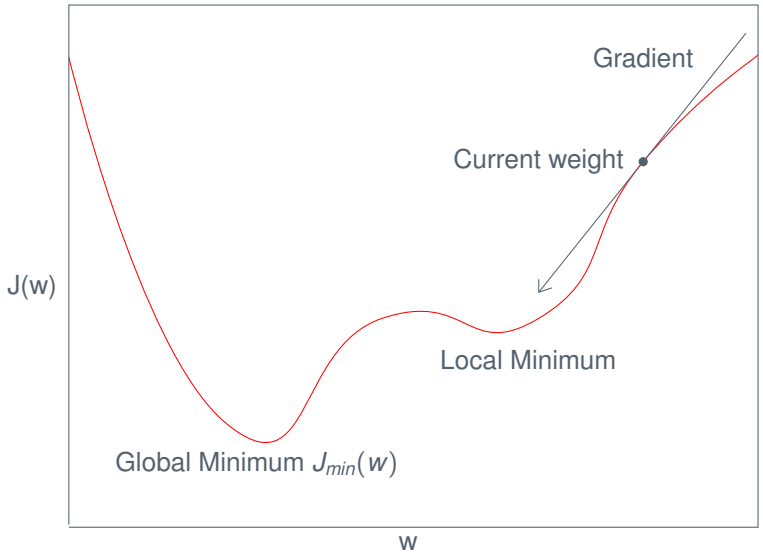
- A set of learnable filters
- Forward pass: convolve each filter across image to produce activation maps
- Local connectivity: Each neuron is connected to a small region of input
- Parameter sharing: Filters useful at different positions

Deep Learning
Training

Gradient Descent

- Minimising a loss function
- Backpropagate loss to update weights & biases

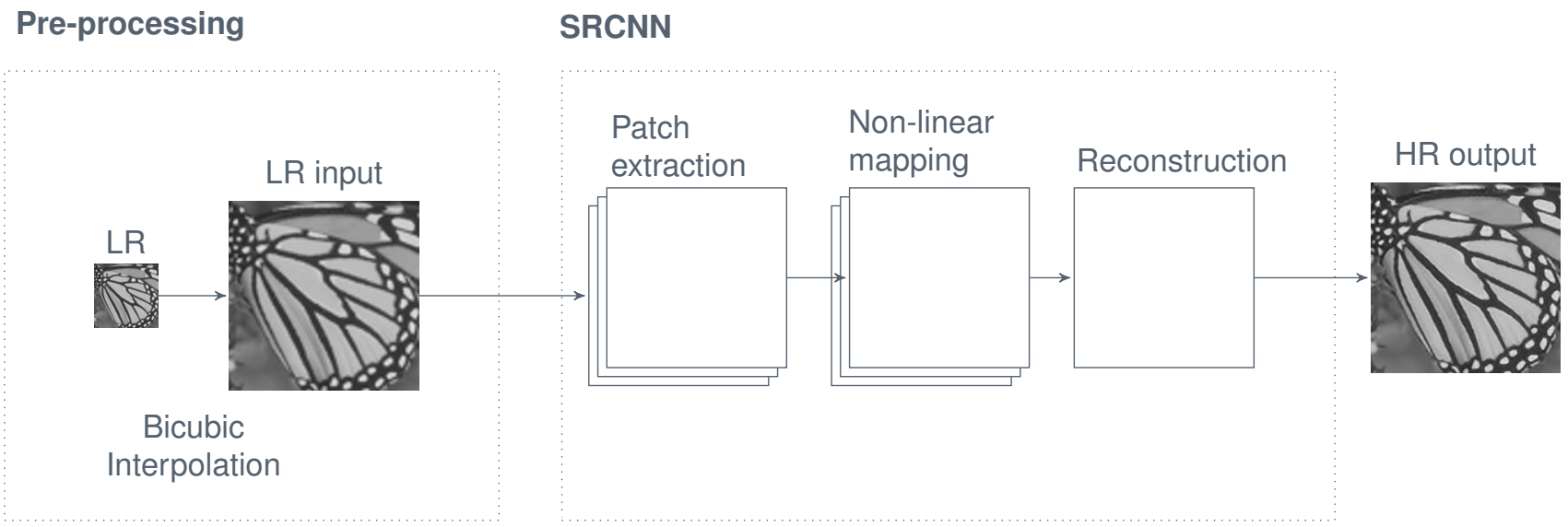




SRCNN
Overview

SRCNN outline

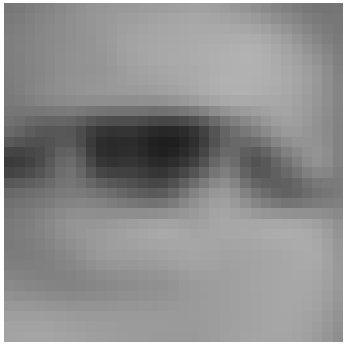
- Pre-processing: Up-scaling via bicubic interpolation
- Three convolutional layers



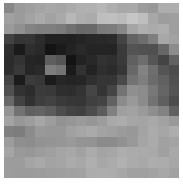
SRCNN
Training

SRCNN Training

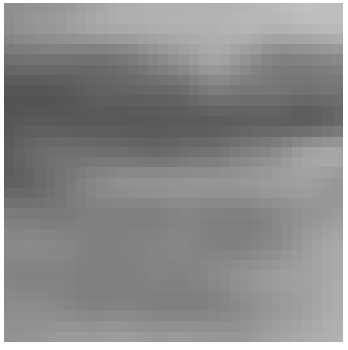
- Supervised learning
- Patch-wise training



(a) LR eye



(b) HR eye

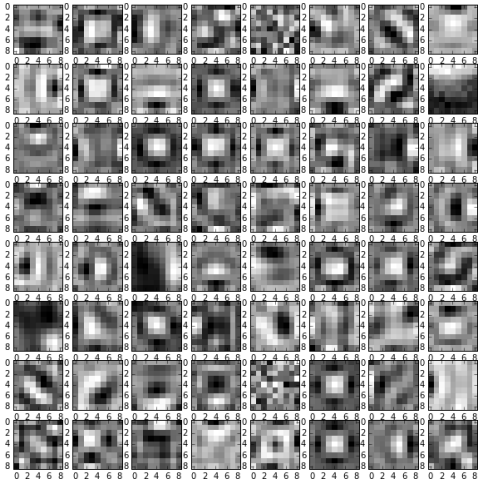


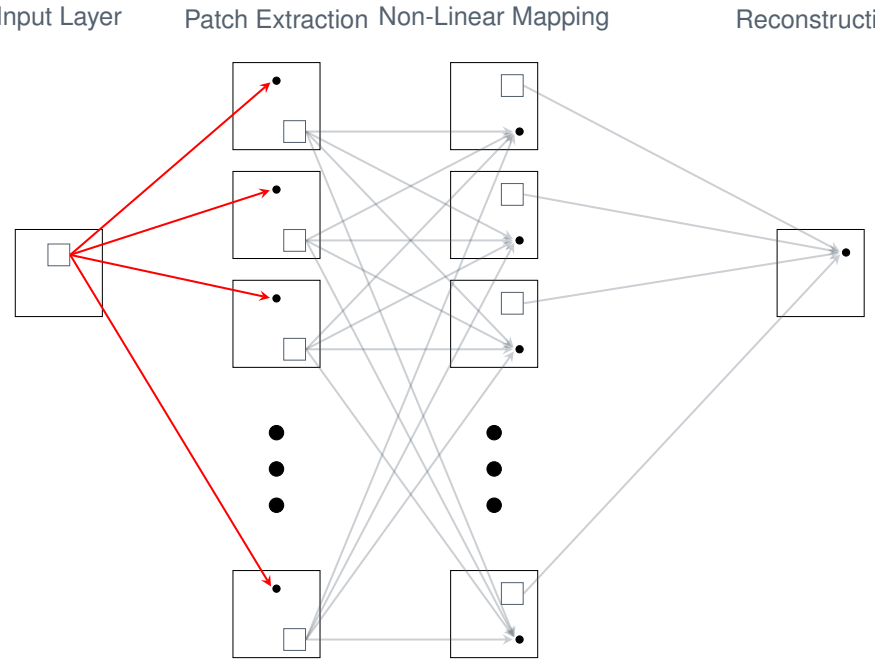
(c) LR mouth

(d) HR mouth

SRCNN
Convolutional Filters

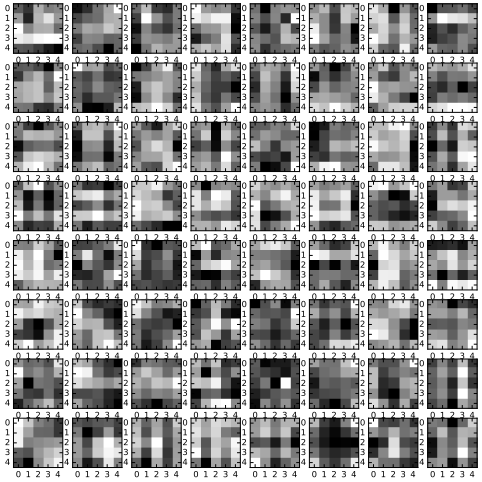
Patch Extraction

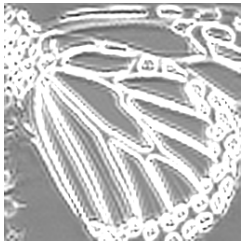


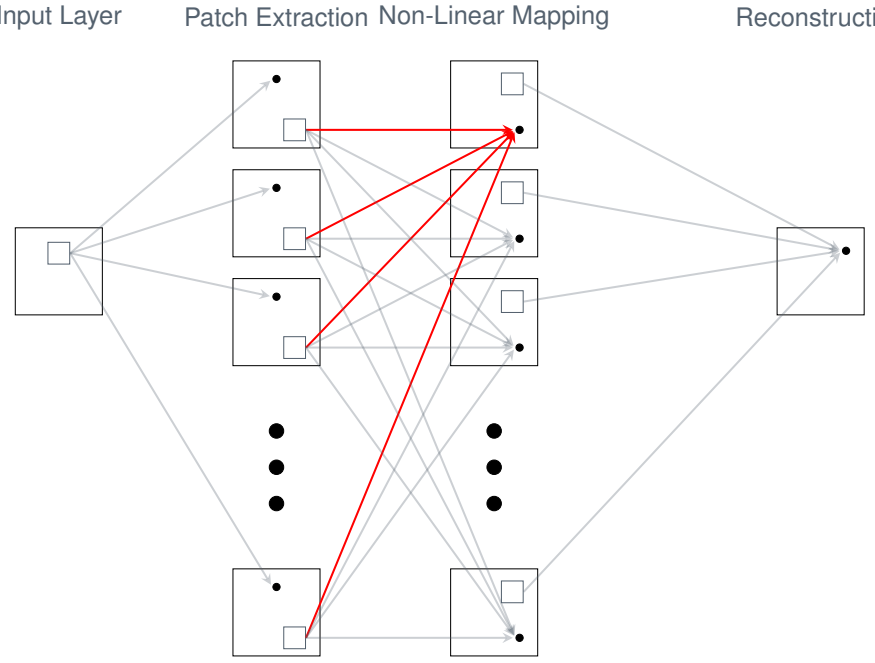


SRCNN
Convolutional Filters

Non-linear Mapping

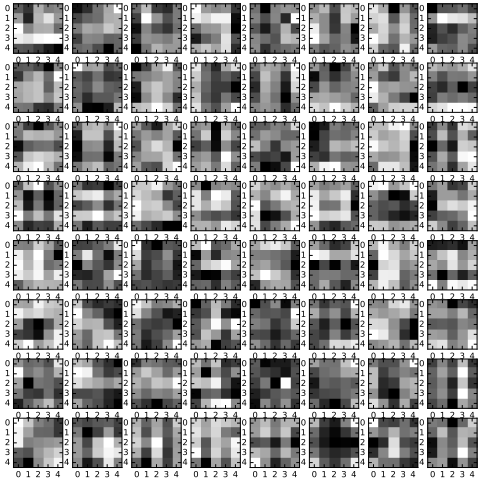


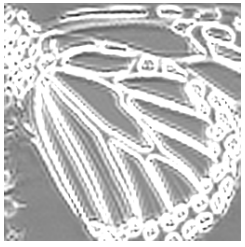


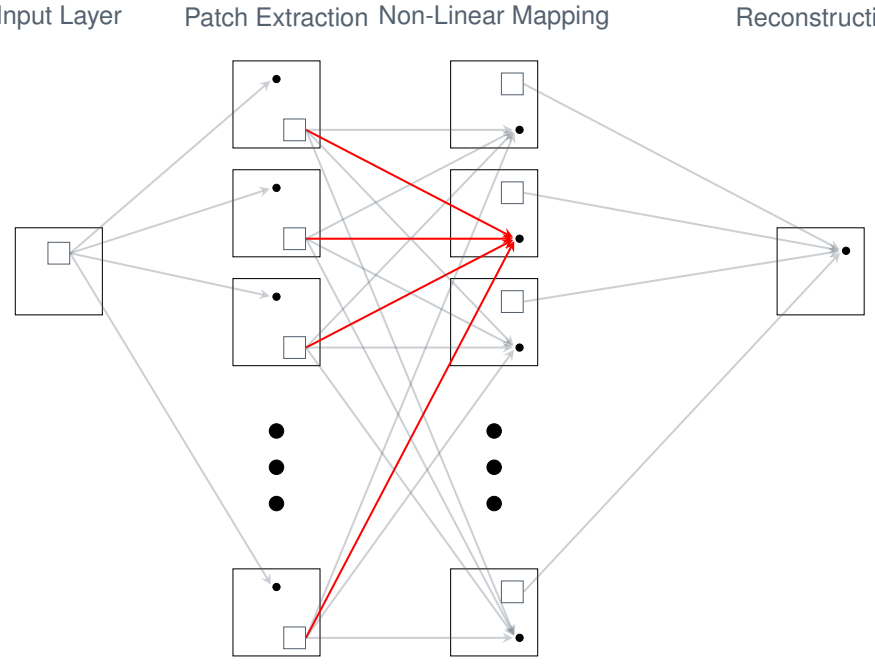


SRCNN
Convolutional Filters

Non-linear Mapping

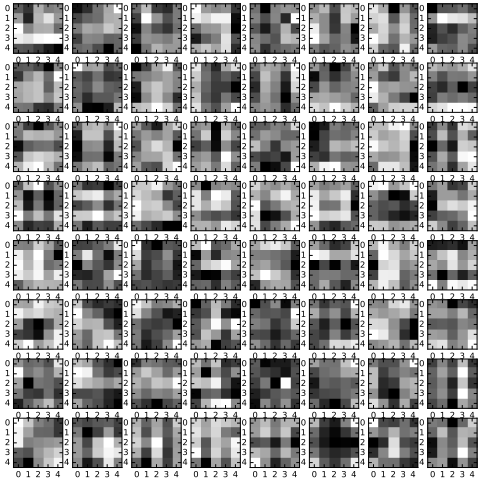


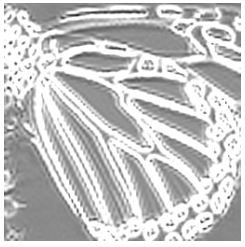


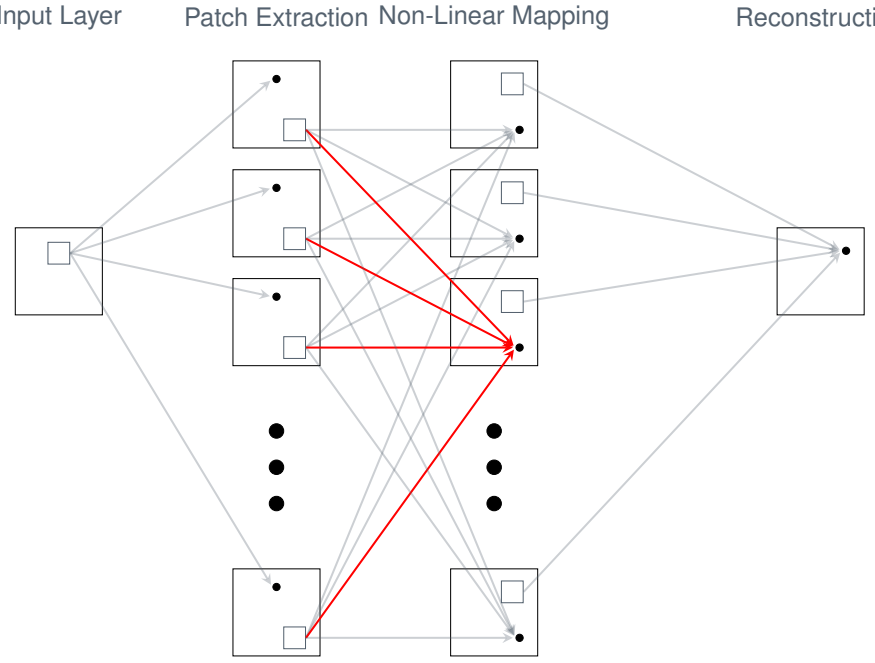


SRCNN
Convolutional Filters

Non-linear Mapping

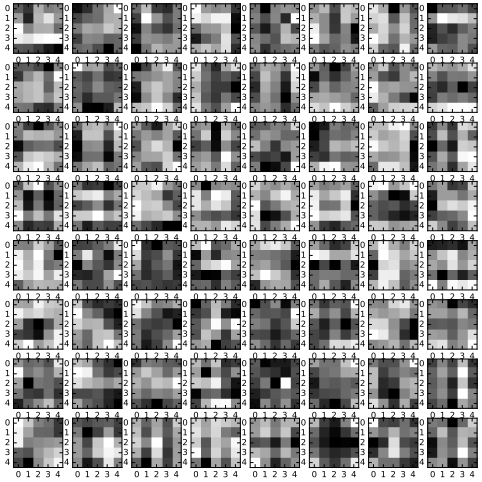


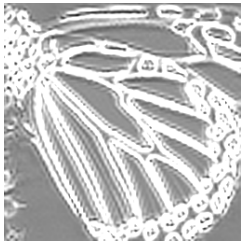


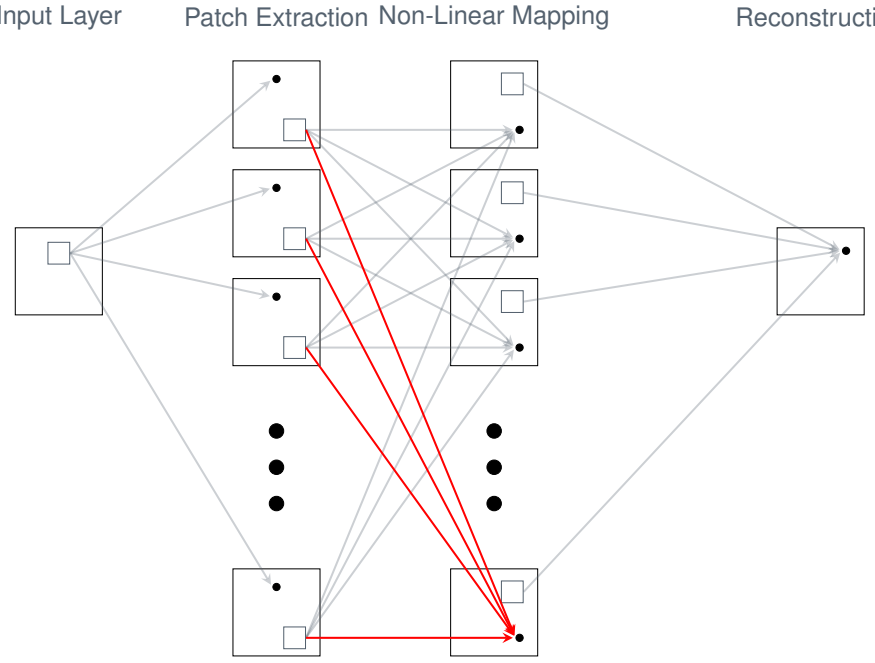


SRCNN
Convolutional Filters

Non-linear Mapping

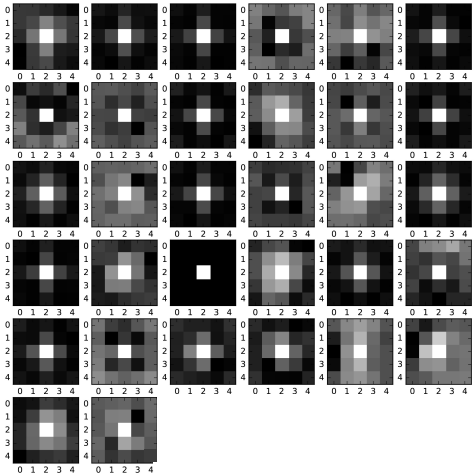


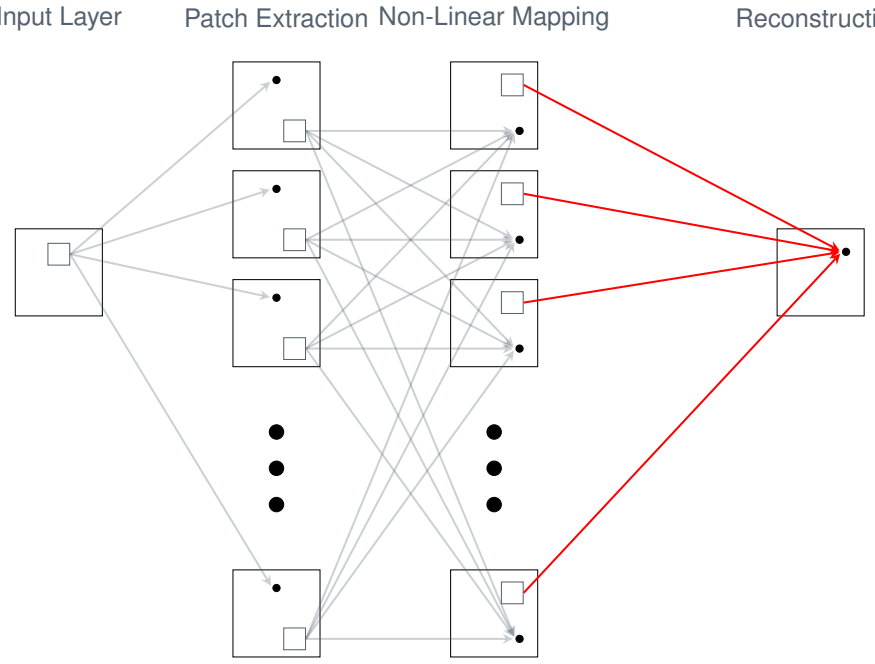




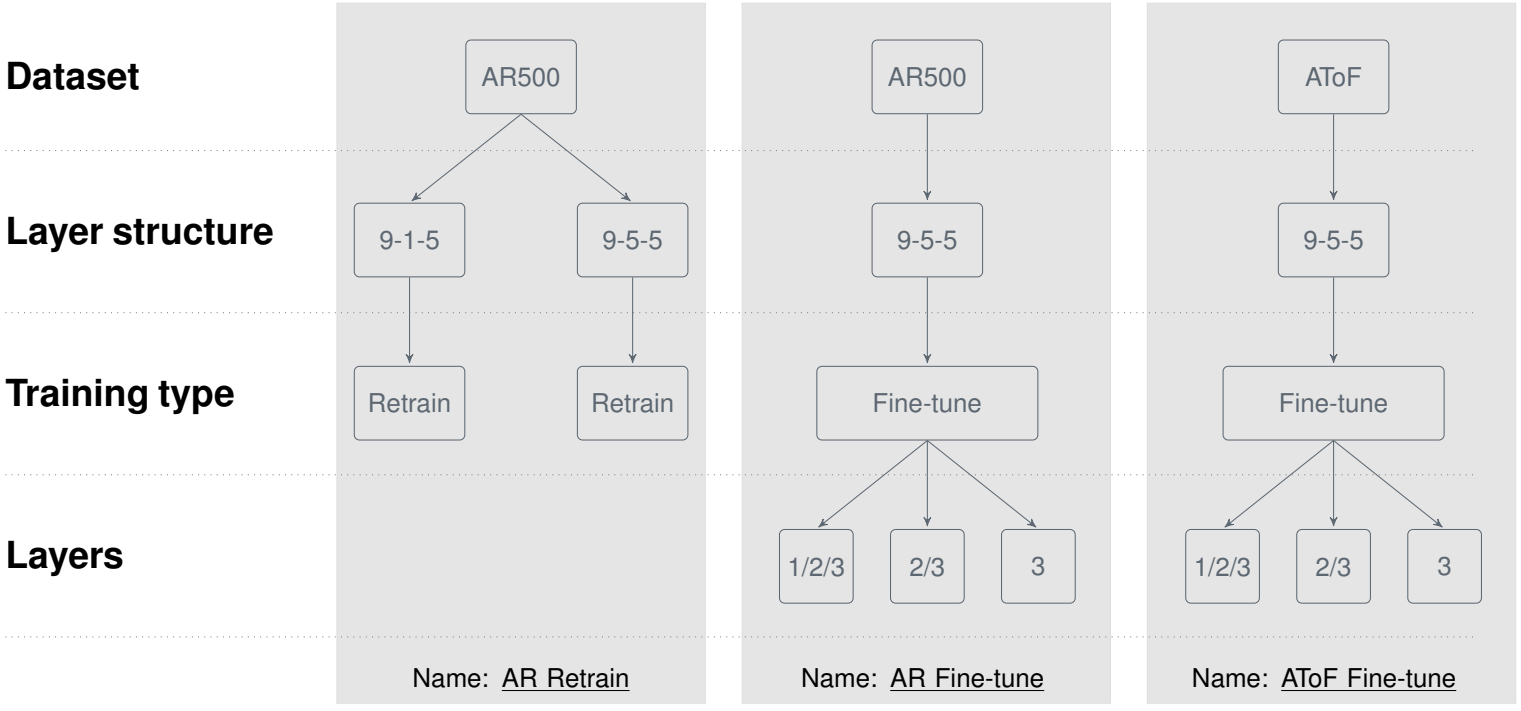
SRCNN
Convolutional Filters

Reconstruction





Research
Overview



Research
Overview

Topics

- Finetuning
- Retraining
- Different types of dataset
- Architectural structure

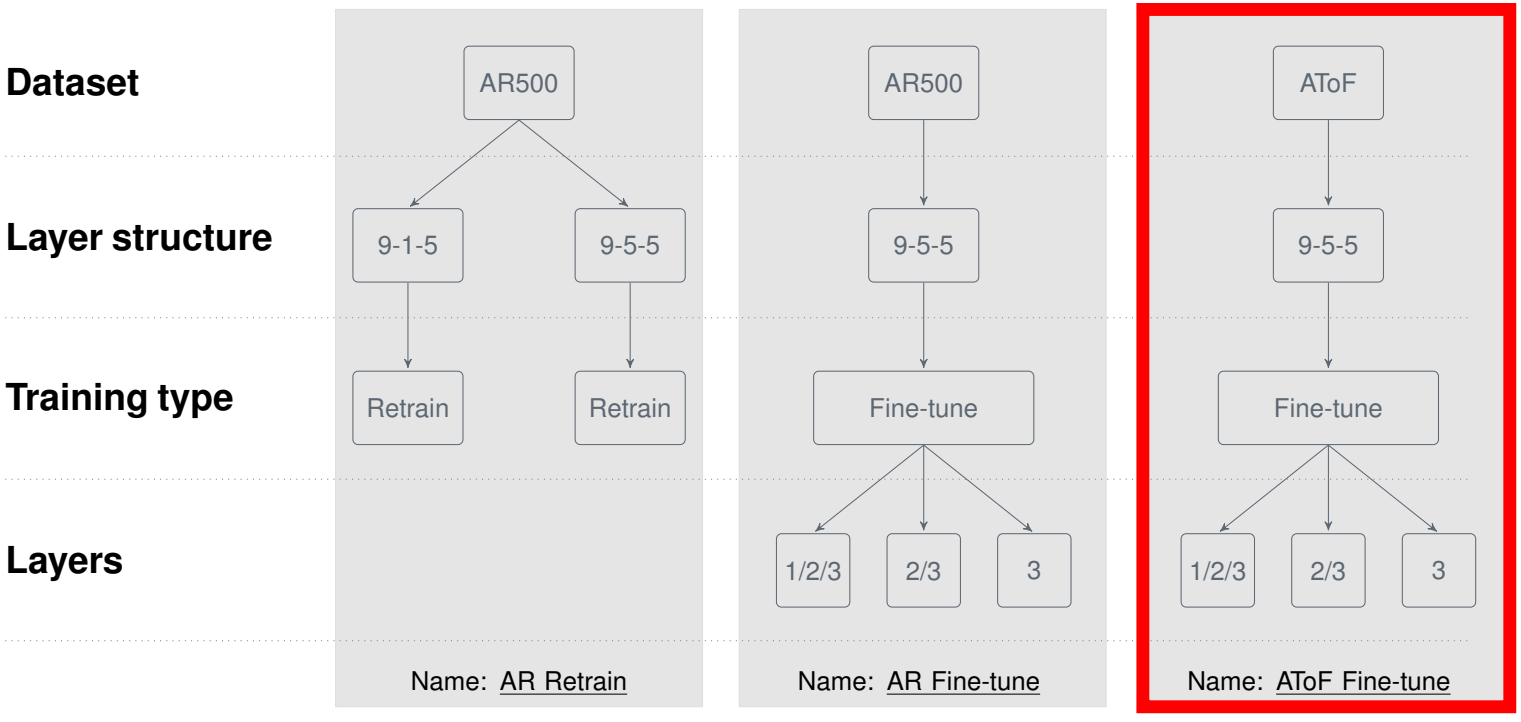
Research
Overview

Topics

- Finetuning
- Retraining
- Different types of dataset
- Architectural structure

Many results!

Research
Overview



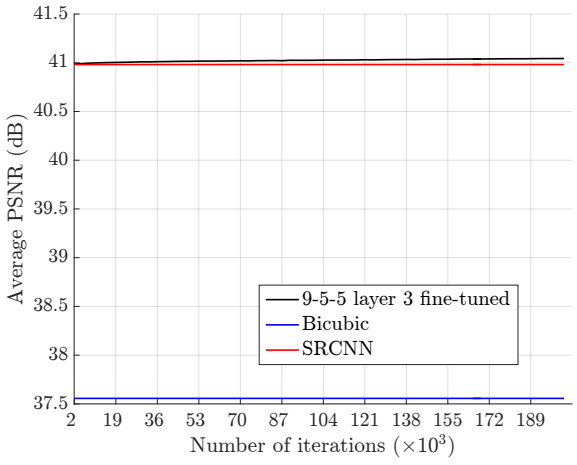
Results
AToF - Finetuning

AToF

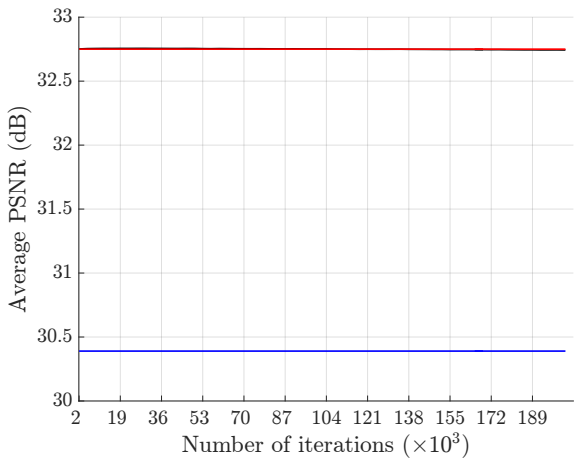
- Not only faces, also background.
- Large variation in face size, resolution and background.

Results

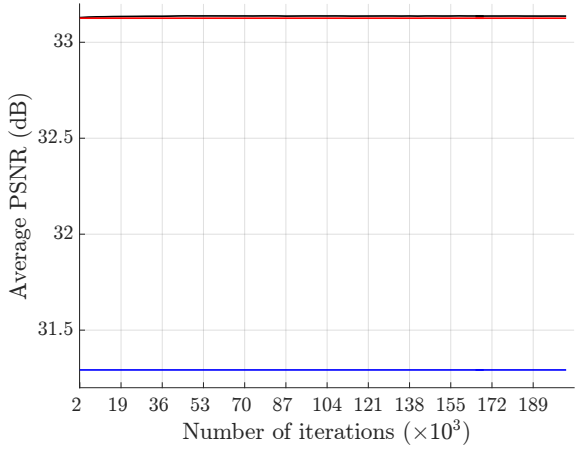
AToF - Finetuning layer 3



AR5



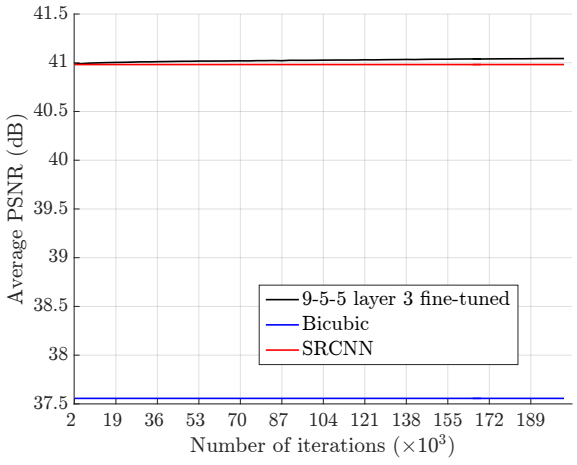
Set5



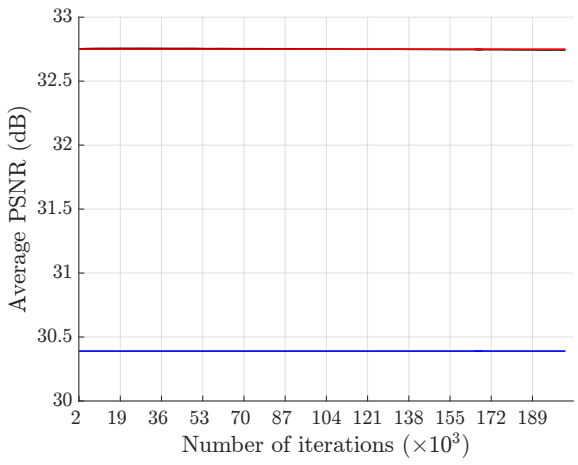
TDRF

Results

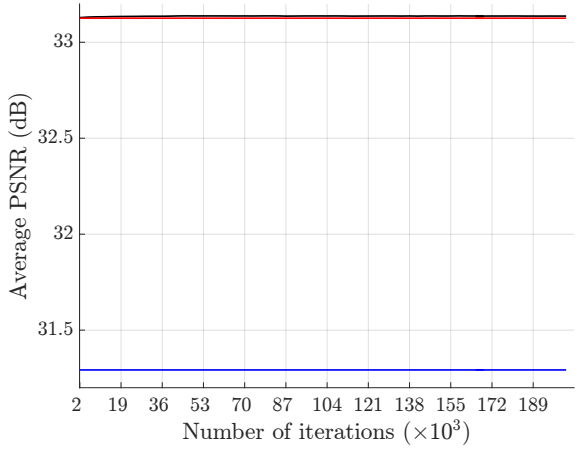
AToF - Finetuning layer 3



AR5



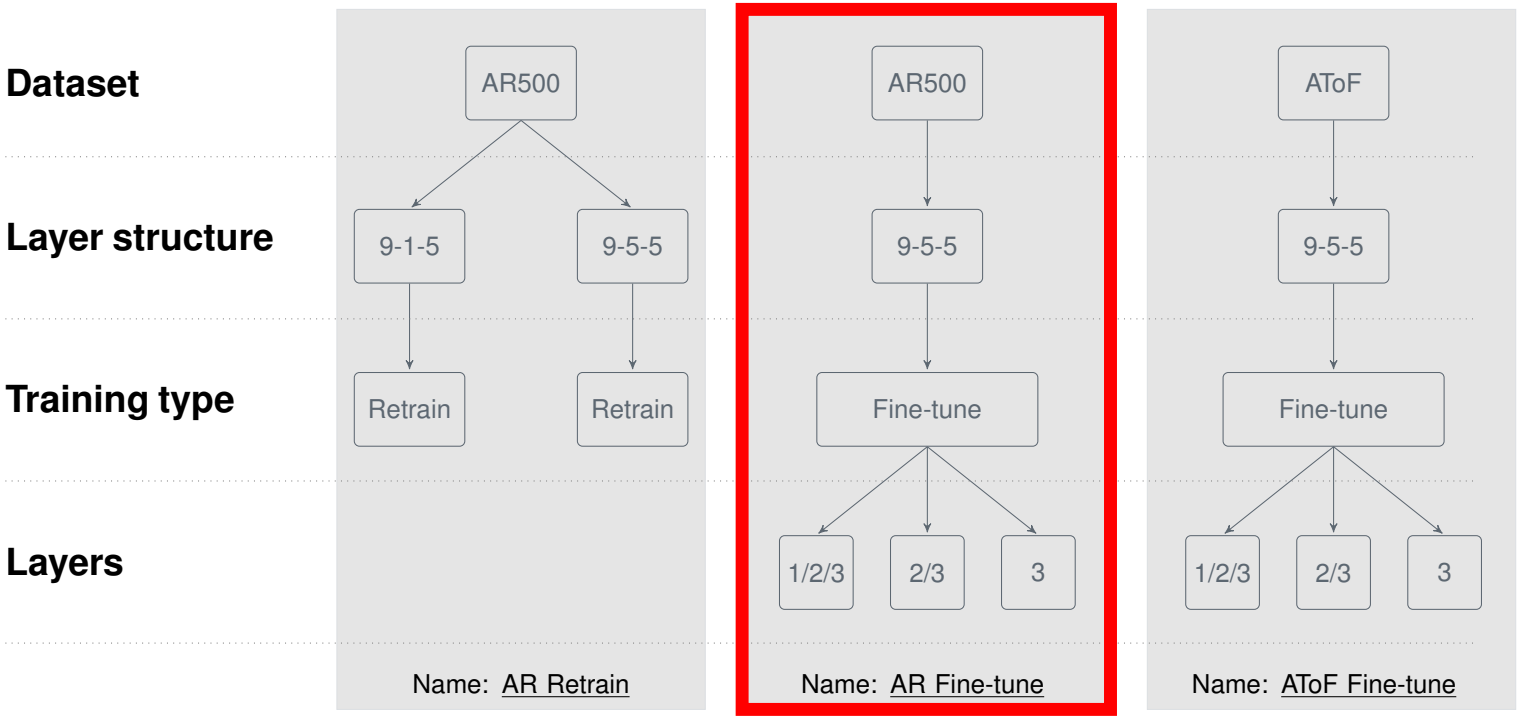
Set5



TDRF

- No significant difference
- No sign on further improvement

Research
Overview



Results

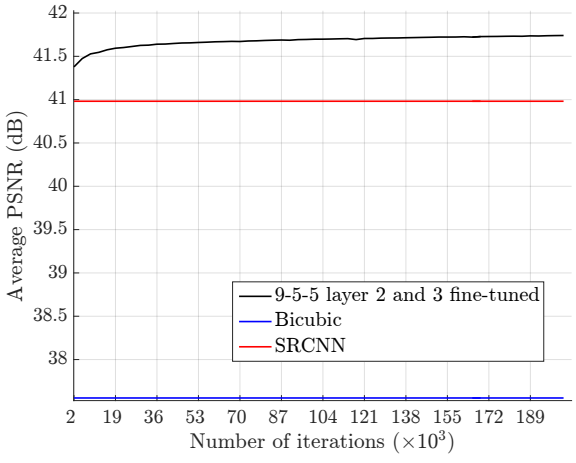
AR500 - Finetuning layer 2 and 3

AR500

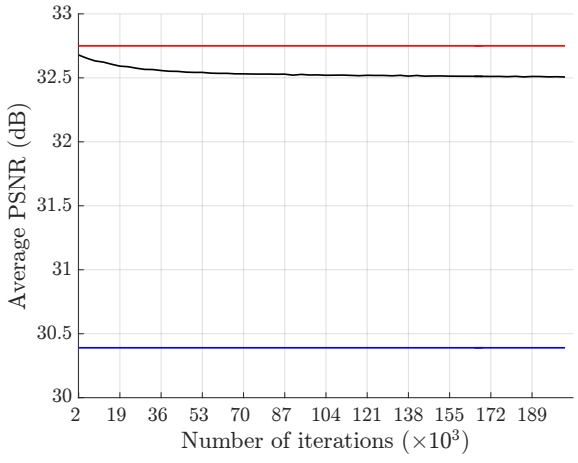
- Same camera and setup.
- No variation in resolution.
- Minor variation in face size and background.

Results

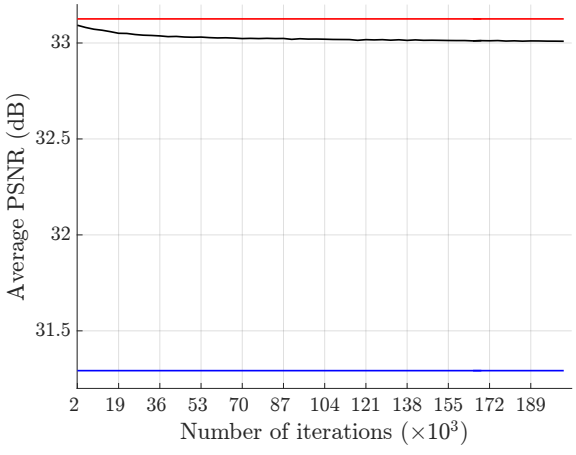
AR500 - Finetuning



AR5

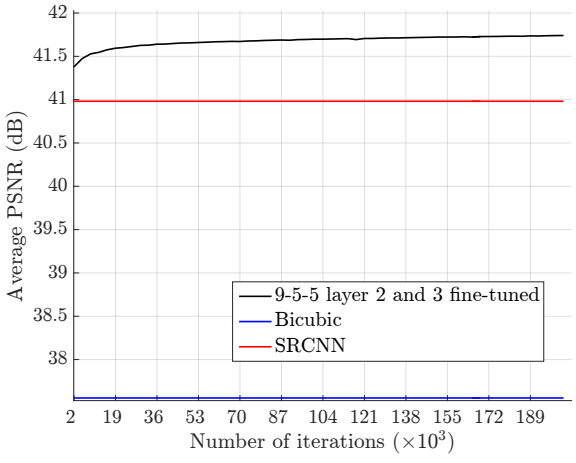


Set5

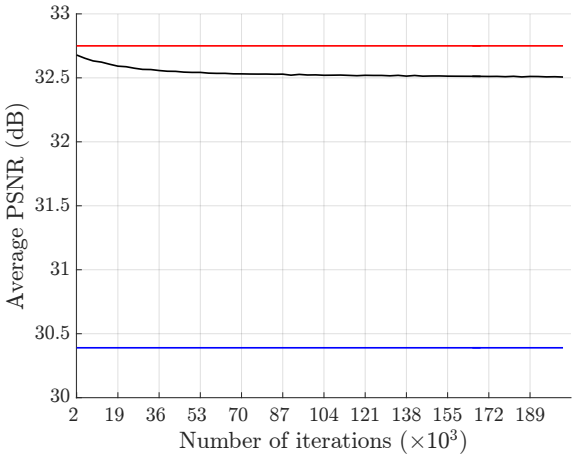


TDRF

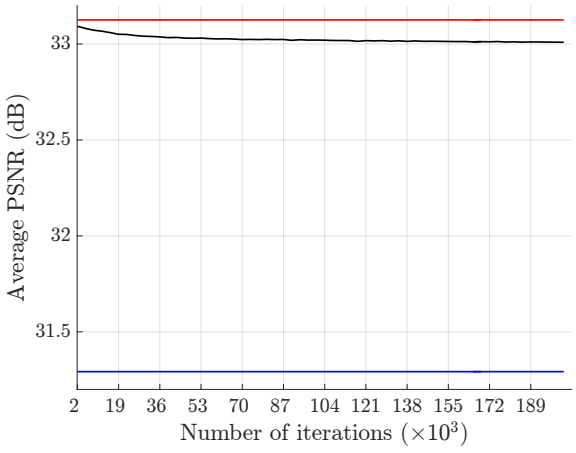
Results
AR500 - Finetuning



AR5



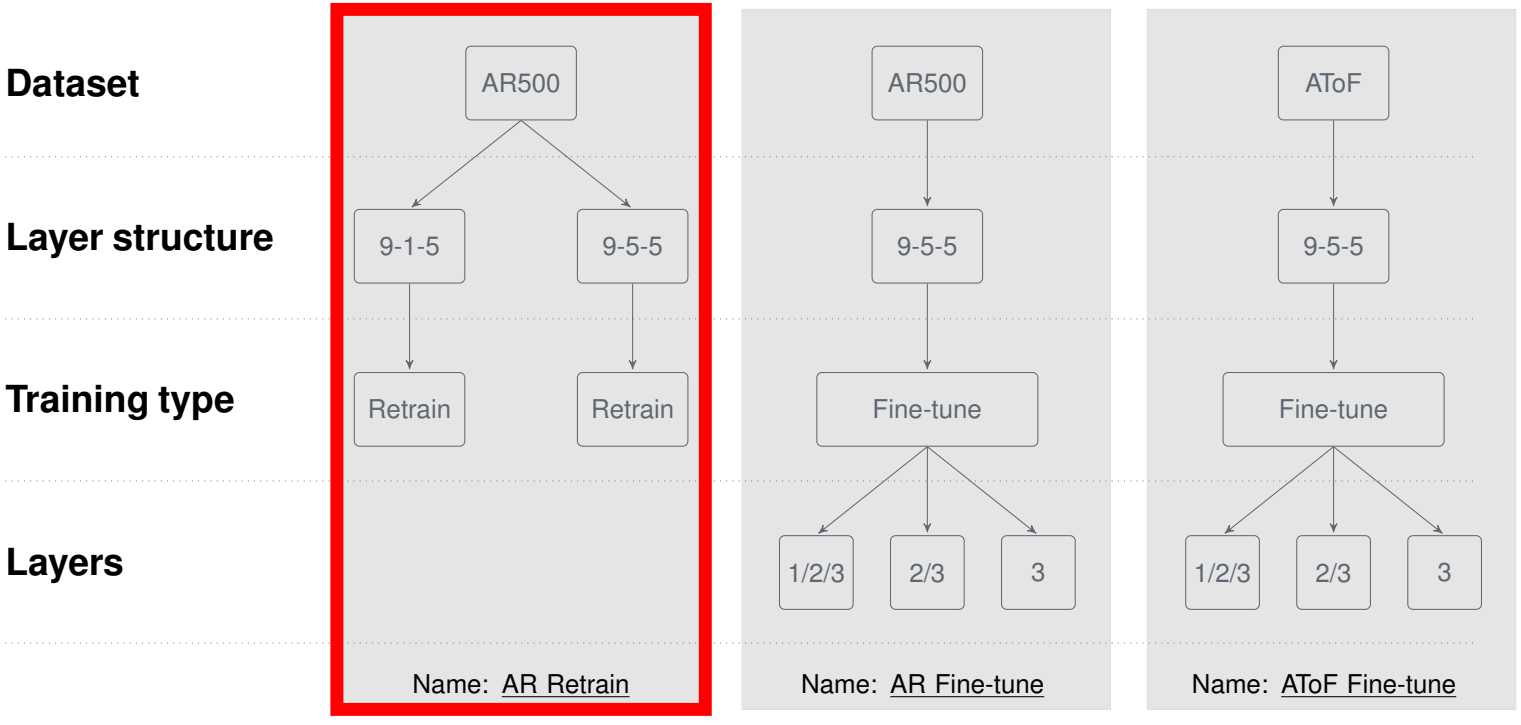
Set5



TDRF

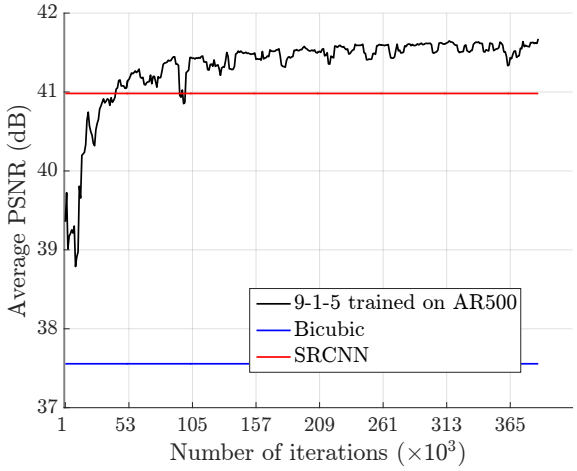
- Worse on images from other environments
- Increase of 0.75dB PSNR on images from similar environments
- May show further improvements with more training

Research
Overview

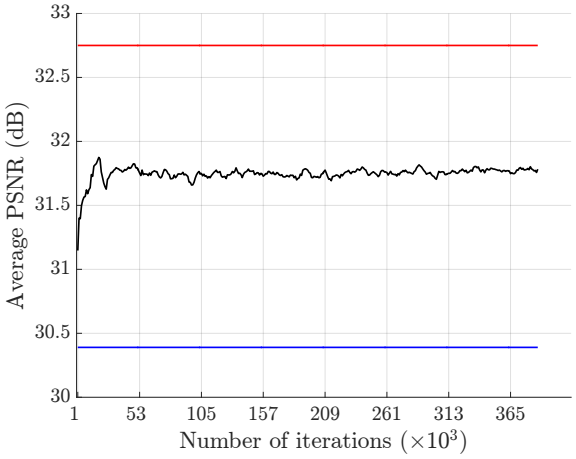


Results

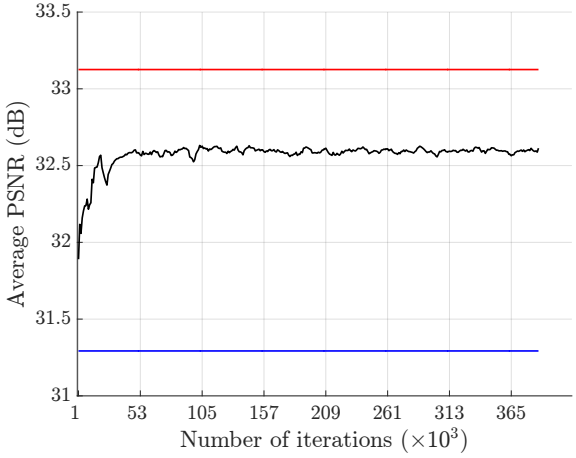
AR500 - Retraining



AR5

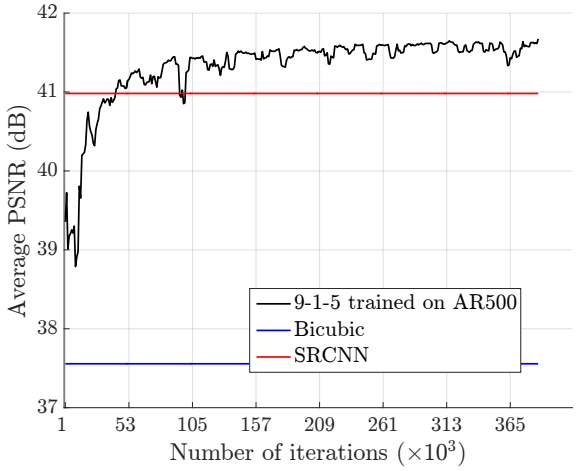


Set5

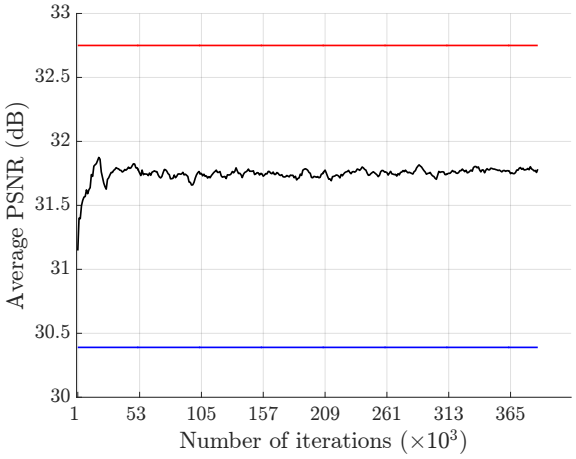


TDRF

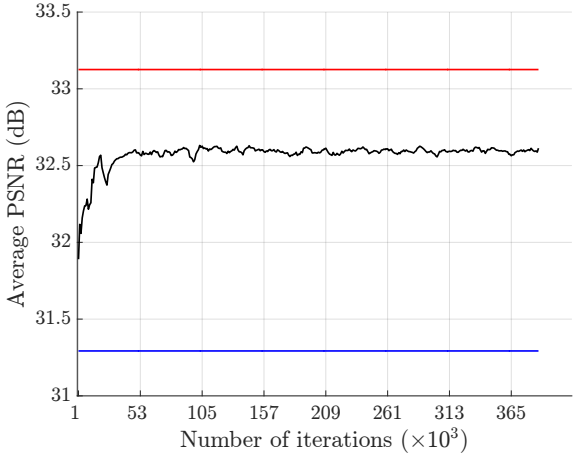
Results
AR500 - Retraining



AR5



Set5



TDRF

- Long training time
- More or less same results as finetuning

Large Face Dataset

New results

Impact of the dataset size in regard to finetuning?

Large Face Dataset
New results

Impact of the dataset size in regard to finetuning?

Full LFD training set (13.300 images):

Dataset	**Bicubic**	**SRCNN**	**SRCNN**$^{2/3}_{AToF}$	**SRCNN**$^{2/3}_{LFD}$
AR5	37.5565/0.9735	40.9816/**0.9814**	**41.1726**/0.9813	39.7784/0.9722
Set5	30.3900/0.8682	32.7500/**0.9090**	**32.7551**/0.9075	32.5721/0.9053
TDRF	31.2926/0.8706	33.1255/**0.9034**	**33.1364**/0.9029	33.0869/0.9026
LFD	32.1229/0.9085	34.1332/0.9338	34.2664/0.9342	**34.5168**/**0.9359**

Large Face Dataset
New results

Impact of the dataset size in regard to finetuning?

Full LFD training set (13.300 images):

Dataset	**Bicubic**	**SRCNN**	**SRCNN**$^{2/3}_{AToF}$	**SRCNN**$^{2/3}_{LFD}$
AR5	37.5565/0.9735	40.9816/**0.9814**	**41.1726**/0.9813	39.7784/0.9722
Set5	30.3900/0.8682	32.7500/**0.9090**	**32.7551**/0.9075	32.5721/0.9053
TDRF	31.2926/0.8706	33.1255/**0.9034**	**33.1364**/0.9029	33.0869/0.9026
LFD	32.1229/0.9085	34.1332/0.9338	34.2664/0.9342	**34.5168**/**0.9359**

LFD-5% training set (666 images):

Dataset	**SRCNN**$^{2/3}_{LFD-5\%}$
AR5	41.1401/0.9813
Set5	32.5356/0.9044
TDRF	33.0602/0.9021
LFD	34.4127/0.9339

Large Face Dataset
New results

Impact of the dataset size in regard to finetuning?

Full LFD training set (13.300 images):

Dataset	**Bicubic**	**SRCNN**	**SRCNN**$^{2/3}_{AToF}$	**SRCNN**$^{2/3}_{LFD}$
AR5	37.5565/0.9735	40.9816/**0.9814**	**41.1726**/0.9813	39.7784/0.9722
Set5	30.3900/0.8682	32.7500/**0.9090**	**32.7551**/0.9075	32.5721/0.9053
TDRF	31.2926/0.8706	33.1255/**0.9034**	**33.1364**/0.9029	33.0869/0.9026
LFD	32.1229/0.9085	34.1332/0.9338	34.2664/0.9342	**34.5168**/**0.9359**

LFD-5% training set (666 images):

Dataset	**SRCNN**$^{2/3}_{LFD-5\%}$
AR5	41.1401/0.9813
Set5	32.5356/0.9044
TDRF	33.0602/0.9021
LFD	34.4127/0.9339

- No significant change in performance

Conclusion

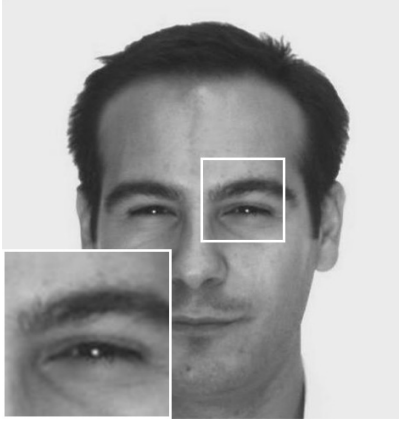
Conclusion

- It is possible to increase the PSNR performance on images from the same environment
- Minor visual difference on the images we have tested on

Conclusion

Conclusion

- It is possible to increase the PSNR performance on images from the same environment
- Minor visual difference on the images we have tested on



Ground truth

Bicubic: 39.04

SRCNN 9-5-5: 42.49

$\text{SRCNN}_{AR}^{2/3}$: 43.41

Conclusion

Conclusion

- It is possible to increase the PSNR performance on images from the same environment
- Minor visual difference on the images we have tested on

Ground truth | Bicubic: 39.04 | SRCNN 9-5-5: 42.49 | $\text{SRCNN}_{AR}^{2/3}$: 43.41

- Difficult to improve performance on all types of face images in one system

Future work

Future work

- Datasets with images of faces only (no background)
- Other image degradations beside downscaling
- Deeper architecture
- Direct investigation of improvements on surveillance images.
  - Train networks on specific cameras and environments

Future work
Datasets

Faces only

Future work
Datasets

Faces only

Degradations

- Blur
- Noise

Future work
Architecture

Deeper architecture

- The research of Kim et al.[5] has shown improvements in performance for deeper networks
- Sophisticated training methods are needed

[5] J. Kim, J. K. Lee, and K. M. Lee, "Accurate Image Super-Resolution Using Very Deep Convolutional Networks," ArXiv e-prints, Nov. 2015.

Optimisation of Convolutional Neural Networks for Super-Resolution on Face Images

June 28, 2016

Andreas Aakerberg, Malte Pedersen, Christoffer B. Rasmussen
16gr843@es.aau.dk

Vision, Graphics and Interactive Systems

Aalborg University

AALBORG UNIVERSITY
DENMARK