Complex & Intelligent Systems (2025) 11:118
https://doi.org/10.1007/s40747-024-01727-2

ORIGINAL ARTICLE

A novel transmission-augmented deep unfolding network with consideration of residual recovery

Zhijie Zhang[1] · Huang Bai[1] · Ljubiša Stanković[2] · Junmei Sun[1] · Xiumei Li[1]

Received: 23 September 2024 / Accepted: 1 December 2024

Abstract

Compressive sensing (CS) has been widely applied in signal processing field, especially for image reconstruction tasks. CS simplifies the sampling and compression procedures, but leaves the difficulty to the nonlinear reconstruction. Traditional CS reconstruction algorithms are usually iterative, having a complete theoretical foundation. Nevertheless, these iterative algorithms suffer from the high computational complexity. The fashionable deep network-based methods can achieve high-precision CS reconstruction with satisfactory speed but are short of theoretical analysis and interpretability. To combine the merits of the above two kinds of CS methods, the deep unfolding networks (DUNs) have been developed. In this paper, a novel DUN named supervised transmission-augmented network (SuperTA-Net) is proposed. Based on the framework of our previous work PIPO-Net, the multi-channel transmission strategy is put forward to reduce the influence of critical information loss between modules and improve the reliability of data. Besides, in order to avoid the issues such as high information redundancy and high computational burden when too many channels are set, the attention based supervision scheme is presented to dynamically adjust the weight of each channel and remove the redundant information. Furthermore, noting the difference between the original image and the output of SuperTA-Net, the reinforcement network is developed, where the main component called residual recovery network (RR-Net) is lightweight and can be added to reinforce all kinds of CS reconstruction networks. Experiments on reconstructing CS images demonstrate the effectiveness of the proposed networks.

Keywords Compressive sensing · Deep unfolding network · Multi-channel transmission · Attention based supervision · Residual recovery · Alternating optimization

Introduction

Compressive sensing (CS) is a signal processing theory that changes the traditional manner of signal sampling and compression, which can significantly improve the sampling efficiency and achieve high-precision reconstruction [1]. CS has been widely applied in signal processing fields such as signal-pixel imaging [2], accelerated magnetic resonance imaging (MRI) [3], video transmission [4] and image codec design [5].

CS aims to recover (or reconstruct) the original high-dimension signal $\boldsymbol{x} \in \mathbb{R}^{n\times 1}$ from its low-dimension measurement $\boldsymbol{y} \in \mathbb{R}^{m\times 1}$ which is obtained by the linear projection $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}$ with $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ the sampling matrix. m/n is usually known as the CS ratio or the sampling ratio. The generic recovery process is formally described as:

$$\arg\min_{\boldsymbol{x}} \frac{1}{2}\|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}\|_2^2 + \omega g(\boldsymbol{x}), \tag{1}$$

Zhijie Zhang and Huang Bai are co-first authors.

✉ Xiumei Li
xiumei_li@hotmail.com

Zhijie Zhang
seghoul1121@163.com

Huang Bai
bh667770@163.com

Ljubiša Stanković
ljubisa@ac.me

Junmei Sun
junmeisun@hznu.edu.cn

[1] School of Information Science and Technology, Hangzhou Normal University, Hangzhou 311121, China

[2] Faculty of Electrical Engineering, University of Montenegro, Podgorica, Montenegro

Published online: 03 January 2025

where $g(\boldsymbol{x})$ is the penalty function on $\boldsymbol{x}$ with ω the penalty parameter [1, 6].

In general, the underdetermined optimization problem (1) with $m \ll n$ leads to extreme difficulties in recovering the original signal from the low-dimension measurement. Many traditional outstanding CS recovery algorithms set $g(\boldsymbol{x})$ to be the ℓ_0-norm or ℓ_1-norm to constrain the sparsity of signal $\boldsymbol{x}$ in some transform domain [7–13]. Such algorithms usually have the advantages of strong convergence and good theoretical analysis. However, solving the ℓ_0-norm or ℓ_1-norm based optimization problem often suffers from the high computational complexity and difficulties on choosing optimal transformation and parameters [14].

Deep learning brings new opportunities to the field of CS, which can solve the CS recovery problem by a data-driven manner. Compared with traditional algorithms, deep learning-based methods can greatly reduce the computational complexity for recovering signals [15–18]. However, deep learning models are composed of repeated fully connected (FC) layers or convolutional layers those are essentially black boxes lacking of good interpretability and theoretical guarantee [19, 20].

As an emerging deep learning model, deep unfolding networks (DUNs) combine the interpretability of the traditional iterative recovery algorithms with the high performance of the deep learning-based methods, which can effectively realize the high-precision reconstruction of the CS tasks [19–30]. Generally, a DUN consists of multiple modules, and each module corresponds to one iteration of the iterative recovery algorithm. The DUNs improve the efficiency of the reconstruction by learning the intrinsic features of the data. Meanwhile, the DUNs retain the interpretability of the iterative recovery algorithms to better optimize the models. Hence, the DUN-based CS reconstruction has become the mainstream to solve the CS problem, and has been widely used to address the signal and image processing issues [19, 20].

In this paper, a novel DUN named supervised transmission-augmented network (SuperTA-Net) is proposed. This network shares a similar architecture with our previous work PIPO-Net [30], where each module in the network can be viewed separately as an optimization problem to be solved with respective penalty function, and the learnable parameters in one module are updated independently from those of other modules within each round of training. Such an architecture makes the network more flexible to find the optimal solutions of the corresponding problems. The main improvements of SuperTA-Net lie in the deep reconstruction modules (DRMs), where a transmission-augmented strategy (also called the multi-channel transmission) is adopted in each DRM to augment the signal reconstruction performance. Besides, noticing the sparse property of the recovery residual between the reconstructed signal and the original signal, an additional reinforcement network is developed to recover the signal residual. The main component of the reinforcement part is termed residual recovery network (RR-Net). Combining the SuperTA-Net with the reinforcement network, we attain the complete SuperTA+RR-Net. Please note that a preliminary version of SuperTA-Net has been published in the 6th International Conference on Data-driven Optimization of Complex Systems. This paper provides a comprehensive expansion of the preliminary version, and offers in-depth exploration and additional insights of the topic. More rigorous theoretical framework and experimental evidence have been appended. In addition, the reinforcement network is totally new from the preliminary version.

The contributions of this paper are three-fold:

- The DUNs for better signal reconstruction by solving the penalty-based CS model are investigated. When unfolding a CS recovery algorithm, critical information may be lost during the transmission between modules, especially in the deep reconstruction stage. A transmission-augmented strategy is proposed to improve the reliability of data by increasing the number of transmission channels.
- The multi-channel transmission plays a significant role in enhancing the signal reconstruction performance. However, it also brings issues such as high redundancy and high computational burden if the channel number is not properly chosen. An attention supervised module is designed in each DRM to dynamically adjust the weights of the channels based on the attention mechanism and remove the redundant information. This results in the SuperTA-Net.
- The residual always exists between the reconstructed signal and the original signal, which may be important in affecting the quality assessment. A reinforcement network is developed to address the residual recovery problem. The main component RR-Net is a lightweight network as the residual is sparse in general. In fact, the reinforcement network is a supplement and can be added to all kinds of CS reconstruction networks. When combined with the SuperTA-Net, the SuperTA+RR-Net is attained.

In addition, extensive experiments on image CS reconstruction are carried out to demonstrate the superior performance of the proposed networks.

The remainder of this paper is arranged as follows. In "Related works", related works on CS recovery methods are introduced. A detailed architecture of SuperTA+RR-Net is described with each part explained in "SuperTA+RR-Net". Experiments are carried out in "Experimental results" to verify the performance of the proposed networks. Some concluding remarks are given in "Conclusion" to end this paper.

Related works

In this section, some CS reconstruction related works will be reviewed, including the traditional iterative algorithms, the classical deep learning-based (also known as data-driven) methods and recent DUN-based ones.

Traditional iterative algorithms

In order to recover signals from CS measurements, traditional strategies utilize sparsity regularization (ℓ_0-norm, ℓ_1-norm or their combination) on the signals in different transform domains to ensure the recovery precision [7–15]. Greedy algorithms such as matching pursuit [7] and orthogonal matching pursuit [8] are popularly adopted methods for solving the ℓ_0-norm constrained optimization problems.

However, the signal recovery precision is hard to be ensured. The convex optimization-based CS recovery methods that replace the ℓ_0-norm constraints with the ℓ_1-norm constraints, usually result in better signal recovery performance [10–12]. Nevertheless, seeking solutions by solving linear programming problems often requires high computational complexity compared to that of the greedy algorithms.

Many traditional iterative algorithms have the advantages of strong convergence and good theoretical analysis by applying structural priors on the signals, such as sparsity and low-rank constraints [31]. However, these algorithms need the development of one or more iterative solvers to enable high-precision signal recovery with empirical regularization construction and parameters selection. In other words, these algorithms inevitably suffer from issues such as high computational costs, poor efficiency, and tedious manual parameters adjustment.

Classical deep learning-based methods

Deep neural networks, with their powerful feature representation capabilities and successful applications in computer vision tasks, have inspired plenty of researchers to investigate the CS recovery methods based on deep learning. Mousavi et al. firstly proposed a method called the stacked denoising auto-encoder (SDA) [15], which aimed to learn effective feature representations of data and to use these features to recover natural images from CS measurements. Inspired by the work of super-resolution, Kulkarni et al. proposed the ReconNet [16], using a convolutional neural network (CNN) architecture for CS reconstruction, which was a non-iterative and extremely fast reconstruction method for block CS imaging. Different from ReconNet, Yao et al. introduced a new architecture called DR^2-Net [17], which was composed of an FC layer-based linear mapping network for preliminary reconstruction and several residual learning blocks to boost the preliminary result. Shi et al. developed a novel image CS framework dubbed CSNet [18], where a sampling network and a reconstruction network were jointly optimized. CSNet utilizes inter-block information effectively and avoids blocking artifacts that are generally encountered in block CS imaging. These deep learning-based methods for CS reconstruction are usually non-iterative, meaning that they could recover signals by a few computations via the neural network. Therefore the time required for reconstructing the signals is greatly reduced.

However, the classical deep learning-based networks are generally trained using the back propagation, which may need substantial labeled data. Besides, these networks adopt end-to-end manners to directly learn the inverse mapping from the measurement domain to the original signal domain, which are essentially black boxes lacking of good interpretability and theoretical guarantee. Moreover, this kind of methods does not take the physical properties of the sampling process and the sampling matrix into consideration, and it is more difficult to learn the inverse mapping without explicit physical guidance.

Recent deep unfolding networks

The DUN is an emerging deep learning model for CS reconstruction. By combining the interpretability of the traditional iterative algorithms with the high performance of the deep learning-based methods, the DUN can achieve fast and accurate CS reconstruction with certain theoretical guarantee.

The iterative process of the traditional iterative algorithm is similar to the stacking process in the neural network and can be realized through a series of stacked trainable neural network modules. The core idea of a DUN is to unfold one CS iterative algorithm into a deep neural network, and each module of the network corresponds to one iteration of the iterative algorithm. It means that the network parameters of the DUN modules are trained with explicit physical guidance. For these reasons, the study of DUNs has attracted growing attention and become one of the mainstreams for inverse imaging tasks [19, 20].

Gregor et al. combined traditional iterative algorithms with deep networks for the first time to provide fast CS reconstruction with high-quality [21]. Their main idea was to train a parameterized non-linear predictor with a specific architecture and a fixed depth to predict the optimal approximation of the sparse representation which could be obtained by traditional iterative algorithms. Inspired by the iterative shrinkage-thresholding algorithm (ISTA) [11] for solving a general ℓ_1-norm CS reconstruction problem, Zhang et al. proposed a structured deep network, dubbed ISTA-Net [22], where an effective strategy to address the proximal mapping associated with the sparsity-inducing regularizer using nonlinear transform was investigated. Yang et al. developed a new DUN named ADMM-CSNet [23], which unfolded

and reformulated the famous alternating direction method of multipliers (ADMM) algorithm [32] to be a deep architecture, where all parameters were discriminatively learned by end-to-end training. Such a network achieved favorable reconstruction accuracy for CS complex-valued MRI applications. A nonconvex optimization inspired multi-scale reconstruction network was introduced by Su et al. for block CS, abbreviated as iPiano-Net [24], by unfolding the inertial proximal algorithm for nonconvex optimization algorithm iPiano [33]. The iPiano-Net was comprised of several cascaded iPiano-like iterations and each iteration contained a block-wise inertial gradient descent and an image-level network-inducing proximal mapping. Zhou et al. proposed a multi-channel deep learning structure for block image CS termed BCSnet [25], which consisted of a channel-specific sampling network for handling block-wise sampling ratio allocation and a unified deep reconstruction network using block-wise approximation and full image-based denoising. BCSnet was the unfolded version of the iterative BCS-SPL algorithm [34]. Zhang et al. proposed the AMP-Net [26], which was established by unfolding the iterative denoising process of the well-known approximate message passing (AMP) algorithm [35], rather than learning regularization terms. Other characteristics of AMP-Net included training the sampling matrix along with the network parameters and integrating deblocking modules to eliminate the blocking artifacts. Different from most DUNs that need to train a specific model for each target sampling matrix, You et al. investigated a controllable arbitrary-sampling network, dubbed COAST [27], to solve CS problems of arbitrary sampling matrices with one single model. In COAST, a random projection augmentation strategy was proposed to promote the generalization ability, and a controllable proximal mapping module and a plug-and-play deblocking strategy were developed to enhance the performance. Shen et al. proposed a novel transformer [36] based hybrid architecture TransCS [28] to achieve high-quality image CS reconstruction at low sampling ratios. A data-driven sampling matrix strategy was adopted in the sampling module and a customized ISTA-based Transformer backbone was integrated with the CNN in the reconstruction module to gain high-performance reconstruction. Existing DUNs unfolded by ℓ_1-norm regularization usually utilize fixed threshold that lacks adaptability for soft-shrinkage operation. Song et al. designed a side-information-aided deep adaptive shrinkage network SODAS-Net [29] to send large volumes of information between adjacent modules, greatly enhancing the network representation capacity. With an effective adaptive soft-shrinkage strategy, SODAS-Net can solve ℓ_1-norm proximal mapping with content-aware thresholds.

SuperTA+RR-Net

This section is devoted to describing the detailed architecture of SuperTA+RR-Net. Each part of this proposed network will be presented with explanation.

Problem formulation

The generic penalty-based CS reconstruction model indicated in (1) is investigated. Instead of learning the prior from data, where handcrafted fixed prior on $g(\boldsymbol{x})$ is set, e.g., ℓ_0-norm or ℓ_1-norm constraint, the learning capacity of the network is wasted. Hence, in this paper, $g(\boldsymbol{x})$ is taken as an adaptive function to be optimized and solved through the network.

By introducing an auxiliary variable $\boldsymbol{z}$, the data fidelity term and the penalty term in (1) can be separated such that:

$$\arg\min_{\boldsymbol{x}} \frac{1}{2}\|\boldsymbol{y}-\boldsymbol{A}\boldsymbol{x}\|_2^2+\omega g(\boldsymbol{z}), \quad \text{s.t.} \quad \boldsymbol{z}=\boldsymbol{x}. \tag{2}$$

Then the following augmented Lagrangian is constructed:

$$\begin{aligned} \mathrm{L}_\rho(\boldsymbol{x},\boldsymbol{z},\boldsymbol{\lambda}) \triangleq\ & \frac{1}{2}\|\boldsymbol{y}-\boldsymbol{A}\boldsymbol{x}\|_2^2+\omega g(\boldsymbol{z})+\boldsymbol{\lambda}^{\mathcal{T}}(\boldsymbol{z}-\boldsymbol{x}) \\ & +\frac{\rho}{2}\|\boldsymbol{z}-\boldsymbol{x}\|_2^2, \end{aligned} \tag{3}$$

where $\boldsymbol{\lambda} \in \mathbb{R}^{n\times 1}$ is the Lagrange multiplier vector, $\mathcal{T}$ is the transpose operator, and ρ is a proper penalty parameter. For the above multivariate optimization problem, the traditional iterative algorithm can alternately optimize the following three sub-problems to reach the solutions:

$$\arg\min_{\boldsymbol{z}} \frac{\omega}{\rho}g(\boldsymbol{z})+\frac{1}{2}\left\|\boldsymbol{z}-\left(\frac{\boldsymbol{\lambda}}{\rho}+\boldsymbol{x}\right)\right\|_2^2, \tag{4}$$

$$\boldsymbol{\lambda} \leftarrow \boldsymbol{\lambda}+\rho(\boldsymbol{z}-\boldsymbol{x}), \tag{5}$$

$$\arg\min_{\boldsymbol{x}} \frac{1}{2}\|\boldsymbol{y}-\boldsymbol{A}\boldsymbol{x}\|_2^2+\boldsymbol{\lambda}^{\mathcal{T}}(\boldsymbol{z}-\boldsymbol{x})+\frac{\rho}{2}\|\boldsymbol{z}-\boldsymbol{x}\|_2^2. \tag{6}$$

Traditionally, (4) is the proximal mapping problem if $g(\boldsymbol{z})=\|\boldsymbol{z}\|_1$. The proximal mapping operator is similar to the denoiser in the network. (5) is essentially a gradient descent updating and (6) has closed-form solutions. Iteratively optimizing (4)–(6) can be unfolded to the deep network and each updating step is viewed as a module in the network.

Complete network architecture

The popular image patch (or block) based network training strategy [22–30] is adopted in this work. The training is carried out for several rounds. For the i-th (round of) training,RR the proposed SuperTA+RR-Net architecture is

shown in Fig. 1, consisting of the SuperTA-Net and the reinforcement network. The SuperTA-Net is the main network to produce the reconstructed image, while the reinforcement network can be treated as a supplement part to enhance the reconstruction performance.

A great many of whole images are split into non-overlapping patches and the patches extracted from the same image should be utilized in the same round of training. Take the famous "Lena" as an example. The whole "Lena" image is first split into $\sqrt{n} \times \sqrt{n}$ non-overlapping patches and each image patch is then vectorized to an $n \times 1$ training (signal) vector. In Fig. 1, the flow of one training vector within the network is demonstrated. The sampling strategy named mean-subtraction sampling (MSS), which was developed in our previous work [30] and has been proved valid for improving the system performance, is adopted to realize the compressive sampling of the vector with its pixel mean removed. The outputs of such a sampling stage include the mean-subtraction measurement and the patch pixel mean. The reconstruction stage of the main network contains one initial reconstruction module (IRM) and K DRMs. The details of the reconstruction stage, especially the DRMs will be introduced in the next subsection. The recovered vectors from the SuperTA-Net can be reshaped into $\sqrt{n} \times \sqrt{n}$ patches for splicing to the whole image. Moreover, each output vector of the SuperTA-Net can also be transmitted to the reinforcement network for further improving the reconstruction accuracy. The supplementary reinforcement network, especially the RR-Net, will be presented later.

Supervised transmission-augmented network

SuperTA-Net is the main network composed of the sampling stage and the reconstruction stage for realizing the image CS reconstruction.

With an input training vector $\boldsymbol{x}_i^*$, the mean-subtraction measurement $\boldsymbol{y}_i$ and the patch pixel mean $\bar{x}_i^*$ can be achieved by the MSS for entering into the reconstruction stage. For convenience, a 1×1 convolutional layer (CONV) is used in the IRM to obtain the initial reconstruction $\boldsymbol{x}_i^{(0)}$. By transmitting $\boldsymbol{x}_i^{(0)}$ pass through K DRMs in sequence, high-precision reconstruction $\boldsymbol{x}_i^{(K)}$ could be achieved.

As can be seen, DRMs play the key roles in the reconstruction. In the proposed network, each DRM mainly consists of four variable update modules followed by a high-frequency complementary (HFC) module. The detailed architecture of the k-th DRM is shown in Fig. 2. The three variable update modules including auxiliary variable denoising module (AVDM), Lagrangian gradient descent module (LGDM) and image patch updating module (IPUM) are designed for solving (4)–(6), respectively. Different from those ADMM-based DUNs, for the proposed network, only the image patch is transferred among DRMs, while the auxiliary variable and the Lagrange multiplier vector are updated within the current DRM. Besides, the attention supervised transmission-augmented module (ASTM) makes the network more distinctive. The HFC is similar to the one that has been developed in our previous work [30], and it is integrated here to compensate the missing high-frequency information and eliminate the blocking artifacts.

To balance the computational complexity and the accuracy, a lightweight denoising network denoted as Net_z is utilized to address (4) for updating the auxiliary variable $\boldsymbol{z}$ without considering the specific form of $g(\boldsymbol{z})$. For computational simplicity, $\omega = \rho$ is always set, where ρ is a learnable parameter in the current DRM. The AVDM architecture is depicted in Fig. 3. The feature extraction part is composed of a 3×3 CONV1, a batch normalization (BN) and a rectified linear unit (ReLU), aiming at extracting the features of the input data. The enhancement part consists of two residual blocks RES1 and RES2 with an ReLU between them. This enhancement part is devoted to reinforcing the features that have been captured in the previous part and it helps to extract the deeper feature representations. The aggregation part contains a BN, a ReLU and a 3×3 CONV2, intending to integrate the obtained features to better reveal the essences of the data. The skip connection facilitates the convergence of the network. Hence, with $\boldsymbol{x}_i^{(k-1)}$ outputted from the $(k-1)$th DRM and $\boldsymbol{\lambda}_{i-1}^{(k)}$ stored in the $(i-1)$th training, $\boldsymbol{z}_i^{(k)}$ can be updated by the AVDM.

Using $\boldsymbol{x}_i^{(k-1)}$, $\boldsymbol{\lambda}_{i-1}^{(k)}$ and the updated $\boldsymbol{z}_i^{(k)}$, $\boldsymbol{\lambda}_i^{(k)}$ is updated via the LGDM. Note that $\boldsymbol{\lambda}_{i-1}^{(k)}$ stored in the memory would be overwritten by the updated $\boldsymbol{\lambda}_i^{(k)}$.

With the obtained $\boldsymbol{z}_i^{(k)}$ and $\boldsymbol{\lambda}_i^{(k)}$, (6) is solved by setting the derivative to zero. This is the result of the IPUM, i.e., $\tilde{\boldsymbol{x}}_i^{(k)}$.

The three modules AVDM, LGDM and IPUM correspond to solving the sub-problems (4)–(6). Such an unfolding procedure sometimes encounters issues such as losing critical information during the transmission between modules. Hence, a transmission-augmented strategy is adopted to improve the reliability of data and that is the ASTM.

Attention supervised transmission-augmented module

In order to reduce the influence of information loss between modules, the multi-channel transmission strategy is adopted. For the $(k-1)$-th DRM, besides $\boldsymbol{x}_i^{(k-1)}$, an $\boldsymbol{x}_i^{(k-1)}$-related information set $\check{\boldsymbol{x}}_i^{(k-1)}$ is also obtained. It should be noted that $\check{\boldsymbol{x}}_i^{(k-1)}$ may be one or more vectors. Combining the result of the IPUM, we get

$$\left\{\hat{\boldsymbol{x}}_i^{(k)}, \check{\boldsymbol{x}}_i^{(k)}\right\} = \tilde{\boldsymbol{x}}_i^{(k)} + \mathcal{M}\left(\tilde{\boldsymbol{x}}_i^{(k)}, \check{\boldsymbol{x}}_i^{(k-1)}\right). \tag{7}$$

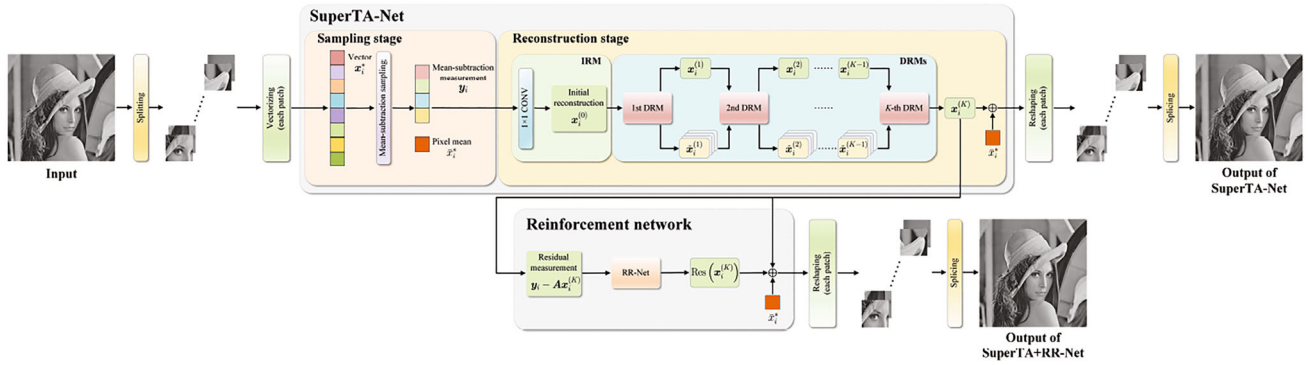

Fig. 1 SuperTA+RR-Net architecture

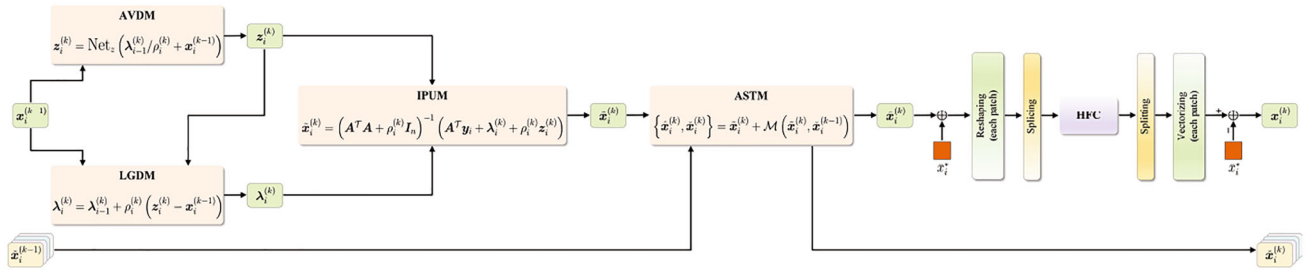

Fig. 2 DRM architecture

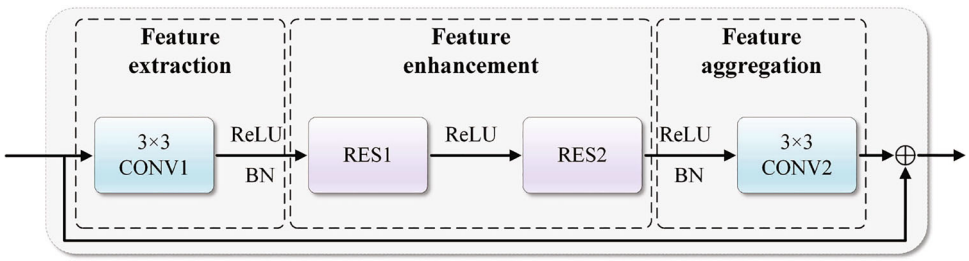

Fig. 3 AVDM architecture

Here $\mathcal{M}(\cdot)$ denotes certain kind of combination operation. It is interesting to note that if

$$\mathcal{M}\left(\tilde{\boldsymbol{x}}_i^{(k)}, \check{\boldsymbol{x}}_i^{(k-1)}\right) = \alpha\left(\tilde{\boldsymbol{x}}_i^{(k)} - \boldsymbol{x}_i^{(k-1)}\right) \tag{8}$$

is set, (7) turns to the famous momentum method which can accelerate the convergence in solving large-scale linear inverse problems [37]. $\mathcal{M}$ is chosen according to the specific requirement. For instance, $\mathcal{M}$ indicated in (8) is a suitable choice for rapid convergence. When one cares more about the information loss, the choices of $\mathcal{M}$ introduced below are preferred.

In order to realize the multi-channel transmission, the following $\mathcal{M}$ is considered:

$$\mathcal{M}\left(\tilde{\boldsymbol{x}}_i^{(k)}, \check{\boldsymbol{x}}_i^{(k-1)}\right) = \mathcal{A}\left(\mathcal{E}\left(\tilde{\boldsymbol{x}}_i^{(k)} \copyright \check{\boldsymbol{x}}_i^{(k-1)}\right)\right). \tag{9}$$

Here, © denotes the channel connection operator, aiming to connect the previous $\check{\boldsymbol{x}}_i^{(k-1)}$ with the current $\tilde{\boldsymbol{x}}_i^{(k)}$ to increase the feature representation ability of the network. $\mathcal{E}(\cdot)$ represents the expansion operation that increases the dimension of the feature map for multi-channel transmission, while $\mathcal{A}(\cdot)$ represents the aggregation operation that decreases the dimension of the feature map for aggregating features. Both $\mathcal{E}$ and $\mathcal{A}$ are implemented by convolutions.

The multi-channel transmission strategy can effectively reduce the influence of information loss between modules and augment the reliability of data. However, it also brings issues such as high information redundancy and high computational burden if too many channels are chosen. To balance the influences of information loss and information redundancy, an attention based supervisor is designed to evaluate the importance of each channel after the expansion operation. The weights of available channels will be dynamically adjusted according to their contributions to the results. The

portion of channels with small weights could be neglected in the current transmission.

Based on the above discussions, the combination operation $\mathcal{M}$ is modified to

$$\mathcal{M}\left(\tilde{\boldsymbol{x}}_i^{(k)}, \check{\boldsymbol{x}}_i^{(k-1)}\right) = \mathcal{A}\left(\mathcal{S}\left(\mathcal{E}\left(\tilde{\boldsymbol{x}}_i^{(k)} \copyright \check{\boldsymbol{x}}_i^{(k-1)}\right)\right)\right) \tag{10}$$

with $\mathcal{S}(\cdot)$ denoting the attention supervised operation. (10) is adopted in the proposed ASTM as indicated in Fig. 2.

The detailed architecture of the ASTM is depicted in Fig. 4. With $\tilde{\boldsymbol{x}}_i^{(k)}$ outputted from the current IPUM and $\check{\boldsymbol{x}}_i^{(k-1)}$ outputted from the previous DRM, the channel connection operation is first executed. Pass the connection result through CONV1 to increase the dimension for information extraction. In order to reduce information redundancy, the multi-channel information is evaluated by the attention based supervisor. By utilizing the global average pooling operation and two FC layers, the weight of each channel is calculated. The channels will be sorted according to their weights (contributions). Only the high-contribution channels are selected and other channels are discarded. The output of the supervisor module is then transmitted to CONV2 for the aggregation operation. Finally, the skip connection is added to facilitate the convergence of the network. It should be noted that the ASTM also produces $\check{\boldsymbol{x}}_i^{(k)}$ for the next DRM.

The combination operation $\mathcal{M}$ indicated in (10) combines the information of the current module with that of the previous DRM. This enhances the connection between adjacent DRMs. Therefore, the ASTM can provide better signal reconstruction performance.

Reinforcement network

As shown in Fig. 1, by transmitting the measurement through the reconstruction stage, the main network SuperTA-Net is sufficient for realizing the image CS reconstruction. However, it is hard to guarantee that all the pixels in the image have been perfectly recovered. Residual always exists between the output of SuperTA-Net and the original image, though it may be very sparse visually.

For one patch with the pixel mean $\bar{x}_i^*$, define the pixel mean vector

$$\bar{\boldsymbol{x}}_i = \left[\bar{x}_i^* \ \bar{x}_i^* \ \cdots \ \bar{x}_i^*\right]^{\mathcal{T}} \in \mathbb{R}^{n\times 1}. \tag{11}$$

The residual measurement can be calculated as

$$\boldsymbol{A}\left(\boldsymbol{x}_i^* - \left(\boldsymbol{x}_i^{(K)} + \bar{\boldsymbol{x}}_i\right)\right) = \boldsymbol{y}_i - \boldsymbol{A}\boldsymbol{x}_i^{(K)}, \tag{12}$$

where $\boldsymbol{y}_i$ is the mean-subtraction measurement, $\boldsymbol{A}$ is the sampling matrix and $\boldsymbol{x}_i^{(K)}$ is the result of the K-th DRM, which are all available in hand. With the residual measurement, we intend to recover the residual $\mathrm{Res}(\boldsymbol{x}_i^{(K)})$ and add it to $\boldsymbol{x}_i^{(K)}$ to further improve the reconstruction. The reinforcement network is designed for this purpose.

The main component of the reinforcement network is the so-called RR-Net which is described in Fig. 5. The RR-Net consists of the initial residual reconstruction and the nonlinear deep reconstruction. A 1×1 convolutional layer is used to initialize the residual which is similar to the IRM of the SuperTA-Net, while the nonlinear deep reconstruction deals with the whole residual image rather than the residual patches which is totally different from the DRMs. Each initialized residual vector should be reshaped to residual patch to splice the whole residual image for entering the nonlinear deep reconstruction. As a lightweight network, the architecture of the nonlinear deep reconstruction is somewhat similar to that of the AVDM depicted in Fig. 3. The feature extraction part is composed of a 3×3 CONV1 and a ReLU activation function, resulting in a high-dimensional feature map. Five residual blocks together with the ReLUs are set in the feature enhancement part to increase the nonlinearity of the network, and this part helps to stepwisely extract the deeper feature representations for the follow-up part. The feature aggregation part integrates the obtained features to generate the residual image. The convergence can be facilitated by fusing the input and the output together through the skip connection. Please note that the output residual image of the nonlinear deep reconstruction should be once again split into patches for further processing.

As long as the main network reconstructs the image well, the residual image is sparse in general. Hence the reinforcement network is lightweight and can be added to reinforce all kinds of CS reconstruction networks.

Loss functions

The reinforcement network is a supplement for enhancing the CS reconstruction performance. So the main network SuperTA-Net could be trained in advance with a loss function $\mathcal{L}_S$ independent from that of the RR-Net. Based on the result of the SuperTA-Net, the RR-Net is then trained with its own loss function $\mathcal{L}_R$. It should be pointed out that the parameters of the SuperTA-Net are kept unchanged during the training of the RR-Net.

For the SuperTA-Net, as introduced in Sect. "Supervised transmission-augmented network", the HFC is integrated in each DRM to compensate the missing high-frequency information and eliminate the blocking artifacts. The design of the HFC is in the same way as that has been developed in our previous work [30]. Based on the investigations in [30], to enforce the HFC to learn the high-frequency information, a two-dimensional discrete wavelet transform could be employed to decompose the whole image into four subband coefficients, and the following wavelet loss function is developed:

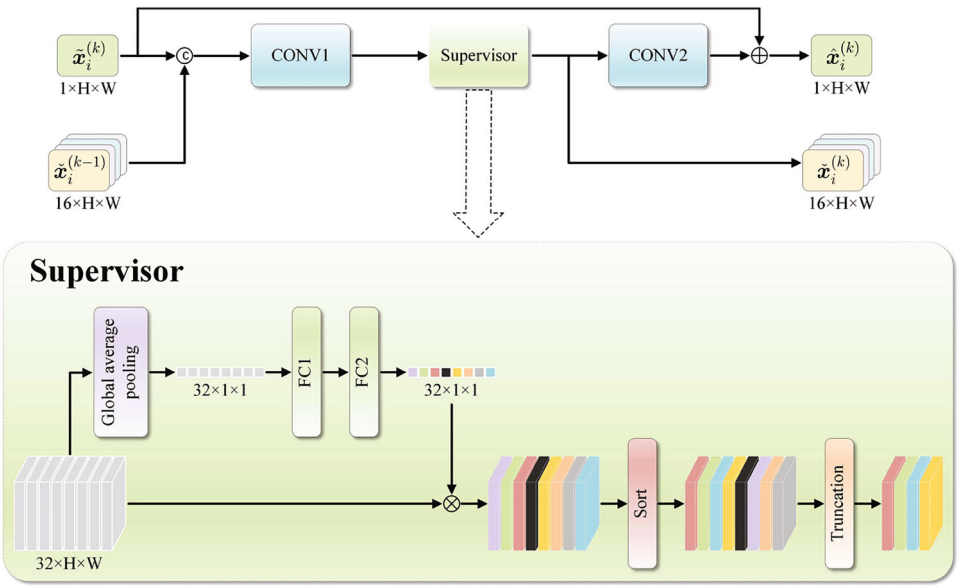

Fig. 4 ASTM architecture

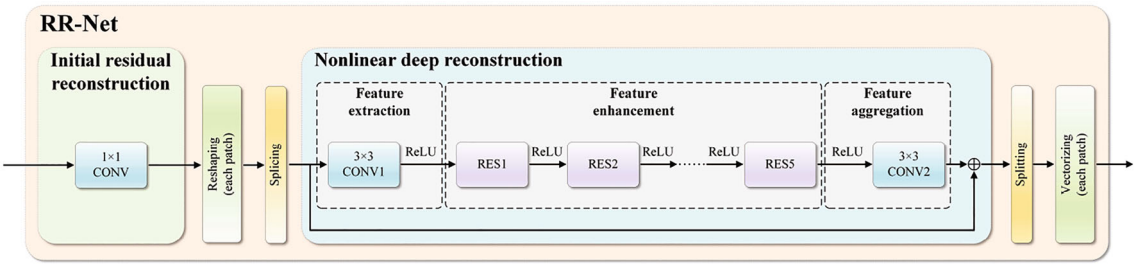

Fig. 5 RR-Net architecture

$$\mathcal{L}_{\mathrm{WT}} \triangleq \frac{1}{NK}\sum_{j=1}^{N}\sum_{k=1}^{K}\left\|\mathcal{W}\left(\boldsymbol{X}_j\right)-\mathcal{W}\left(\mathcal{H}\left(\hat{\boldsymbol{X}}_j^{(k)}\right)\right)\right\|_F^2, \tag{13}$$

where $\mathcal{W}(\cdot)$ denotes the Haar wavelet decomposition operation, $\mathcal{H}(\cdot)$ is the HFC operation, $\|\cdot\|_F$ is the Frobenius norm, K is the number of HFC modules (also the number of DRMs), N is the number of whole images chosen for the i-th training. $\boldsymbol{X}_j$ is one original whole image and $\hat{\boldsymbol{X}}_j^{(k)}$ is the corresponding whole image as the HFC input in the kth DRM.

Recalling the definition of the commonly adopted mean square error (MSE) loss function:

$$\mathcal{L}_{\mathrm{MSE}} \triangleq \frac{1}{\mathrm{numel}\left(\boldsymbol{X}_j\right)N}\sum_{j=1}^{N}\left\|\boldsymbol{X}_j-\dot{\boldsymbol{X}}_j\right\|_F^2, \tag{14}$$

with numel$(\cdot)$ counting the number of pixels in one image and $\dot{\boldsymbol{X}}_j$ being the corresponding output image of the network, the total loss function of the SuperTA-Net is presented as

$$\mathcal{L}_S = \mathcal{L}_{\mathrm{MSE}} + \gamma\mathcal{L}_{\mathrm{WT}}, \tag{15}$$

where γ is the regularization parameter and γ=0.1 leads to satisfactory results.

As the RR-Net is relatively lightweight, the following loss function is utilized for training:

$$\mathcal{L}_R \triangleq \frac{1}{\mathrm{numel}\left(\mathcal{X}_j\right)N}\sum_{j=1}^{N}\left\|\mathcal{X}_j-\dot{\mathcal{X}}_j\right\|_F^2, \tag{16}$$

where $\mathcal{X}_j$ is the difference between the original input image and the output of the SuperTA-Net, and $\dot{\mathcal{X}}_j$ is the corresponding recovered whole residual image.

Experimental results

In this section, the proposed networks SuperTA-Net and SuperTA+RR-Net are both applied for the image CS recon-

struction. Seven recent deep unfolding network-based methods, iPiano-Net [24], BCSnet [25], AMP-Net [26], COAST [27], TransCS [28], SODAS-Net [29] and PIPO-Net [30], are chosen for comparison.

Experimental settings

The experimental settings are similar to that of the AMP-Net. The BSDS500 dataset [38] which contains 500 colorful images are utilized. This dataset is divided into a training set of 200 images, a validation set of 100 images, and a testing set of 200 images. The colourful images are processed in the YCbCr space and the luminance components are used for training and validation. In the training set, 448 sub-images with the size 99×99 are randomly extracted from the luminance component of each image, resulting in total 89,600 sub-images. The 99×99 sub-images are used as the input images to the networks and this leads to numel $\left(\boldsymbol{X}_j\right) = 99 \times 99 = 9801$ in (14). The size of an image patch is set as 33×33, hence $n = 33 \times 33 = 1089$. It is clear that for each sub-image, 9 non-overlapping patches can be extracted. The Adam optimizer [39] with a learning rate of 0.0001 and a batch size of 256 (i.e., $N = 256$) is used to train the network, with 150 epochs for each CS ratio. For testing, besides the BSDS500 testing set, two popular benchmark datasets, Set11 [16] and BSD68 [40], are utilized. The peak signal-to-noise ratio (PSNR) in dB and the structural similarity (SSIM) [41] are chosen to evaluate the performance. The experiments are conducted on a deep learning workstation with an Intel i7-13700 CPU and GTX4080 GPU.

Initialize the sampling matrix $\boldsymbol{A}$ as a random Gaussian matrix with its rows orthogonalized. The same sampling matrix optimization strategy is adopted in the proposed network as that in the AMP-Net, where the sampling matrix is jointly trained with other network parameters to improve the reconstruction accuracy. Considering the trade-off between network complexity and reconstruction performance, the default number of DRMs of the SuperTA-Net is set to $K = 9$, while traditional iterative optimization algorithms usually require $K > 100$.

Ablation studies

Firstly, several groups of ablation studies are carried out to test the effects of the three main contributions of this paper, including the multi-channel transmission-augmented strategy, the attention based supervision scheme and the reinforcement network. Each group of the ablation studies is conducted on Set11 with three CS ratios 1%, 10% and 25%.

The first group of ablation studies is conducted to verify the importance of the combination operation $\mathcal{M}$ indicated in (7) which is prepared for the multi-channel transmission. As mentioned in Sect. "Attention supervised transmission-augmented module", $\mathcal{M}$ is chosen according to the specific requirement. Four $\mathcal{M}$ choosing schemes are tested and the results are summarized in Table 1.

$\mathcal{M} = 0$ means that no combination operation is added, i.e., there exists no transmission augmentation between modules. This can be viewed as the benchmark in this test. Equation (8) refers to the conventional momentum acceleration method. Such a choice makes the problem to converge to a relative better solution, e.g., improving the PSNR by 0.17 dB with 10% CS ratio. However, the momentum acceleration method could not avoid the information loss. In order to demonstrate the influence of the information compensation, a rough multi-channel (32 channels) transmission scheme without supervision, i.e., (9) is adopted. This choice brings the improvements of the PSNR by 0.21 dB, 0.32 dB and 0.28 dB with 1%, 10% and 25% CS ratio, respectively. Nevertheless, this rough augmentation scheme also leads to high information redundancy and high computational complexity. Equation (10) is exactly the supervised scheme adopted in the proposed ASTM. With the most important 50% of total 32 channels retained, the scheme (10) reaches much better results in all the three CS ratio cases, and the superiority is clearer when low CS ratio is set, e.g., 1%.

Table 2 summarizes the results of the second group of ablation studies to assess the redundancy of the multi-channel transmission when scheme (9) is chosen. The number of channels is set to 1, 8, 16 and 32, respectively. In all the three CS ratio cases, the scheme with 16 channels always performs best. This leads to a conclusion that the number of transmission channels should be constrained to avoid the information redundancy.

As observed in Table 2, the redundant information is useless and even harmful for the network. Therefore, we design an attention based supervisor to determine which channels are more important to be selected with the rest to be discarded. 32 channels are initially set. With the percentage given, the supervisor is used to select the channels for transmission. Four cases, 3%, 25%, 50% and 100% of the 32 channels, are tested for comparison. The PSNR (dB)/SSIM results are presented in Table 3. It can be seen that the scheme with 50% of the 32 channels, i.e., 16 channels selected for transmission is preferable.

Please note that (10) is very different from (9). For scheme (9), the total channel number is varying for each group of test, while for scheme (10), the total channel number is fixed as 32 and different percentages of the 32 channels are selected for transmission. Besides, for the 32 channels case where the channel numbers chosen for transmission are the same between (9) and (10), the network with (10) still performs better than that with (9) because the supervisor dynamically adjusts the weights of the channels according to their contributions and this is the main characteristic of the proposed attention based supervision scheme.

Table 1 Ablation studies on assessing the performance of different combination operations on Set11 (the best results are highlighted in bold)

Combination operations	1%	10%	25%
$\mathcal{M}=0$	22.52/0.5844	29.34/0.8580	34.83/0.9433
$\mathcal{M}=\alpha\left(\tilde{\boldsymbol{x}}_i^{(k)}-\boldsymbol{x}_i^{(k-1)}\right)$, i.e., (8)	22.60/0.5881	29.51/0.8783	35.04/0.9498
$\mathcal{M}=\mathcal{A}\left(\mathcal{E}\left(\tilde{\boldsymbol{x}}_i^{(k)}©\check{\boldsymbol{x}}_i^{(k-1)}\right)\right)$, i.e., (9)	22.73/0.5901	29.66/0.8775	35.11/0.9502
$\mathcal{M}=\mathcal{A}\left(\mathcal{S}\left(\mathcal{E}\left(\tilde{\boldsymbol{x}}_i^{(k)}©\check{\boldsymbol{x}}_i^{(k-1)}\right)\right)\right)$, i.e., (10)	**23.73/0.6052**	**30.53/0.9113**	**35.61/0.9594**

Table 2 Ablation studies on assessing the redundancy of the multi-channel transmission with (9) on Set11 (the best results are highlighted in bold)

Number of channels	1%	10%	25%
1	22.66/0.5818	29.20/0.8515	34.84/0.9477
8	22.75/0.5922	29.82/0.8831	35.02/0.9494
16	**22.93/0.6037**	**29.94/0.8972**	**35.22/0.9503**
32	22.73/0.5901	29.66/0.8775	35.11/0.9502

Table 3 Ablation studies on assessing the performance of the ASTM channel selections on Set11 (the best results are highlighted in bold)

Percentage of channels	1%	10%	25%
3% (1 channel)	22.97/0.5723	29.21/0.8498	34.88/0.9482
25% (8 channels)	23.55/0.5923	30.11/0.8857	35.47/0.9550
50% (16 channels)	**23.73/0.6052**	**30.53/0.9113**	**35.61/0.9594**
100% (32 channels)	23.57/0.5919	30.01/0.8664	35.35/0.9533

Table 4 Ablation studies on assessing the effectiveness of the RR-Net on Set11

Networks	1%		10%		25%	
	w/o RR-Net	with RR-Net	w/o RR-Net	with RR-Net	w/o RR-Net	with RR-Net
ReconNet [16]	20.09/0.5201	20.14/0.5311	24.06/0.7782	24.13/0.7903	25.60/0.7613	25.72/0.7528
ISTA-Net [22]	17.34/0.4403	17.36/0.4406	28.34/0.7341	28.42/0.7355	32.57/0.9176	32.87/0.9282
AMP-Net [26]	20.21/0.5581	20.36/0.5588	29.13/0.8610	29.52/0.8809	34.64/0.9488	34.88/0.9496
SuperTA-Net	22.29/0.5872	23.73/0.6052	30.37/0.9084	30.53/0.9113	35.27/0.9519	35.61/0.9594

Table 5 PSNR (dB)/SSIM results of the nine DUN methods on BSDS500 testing set (the top and the second-best results are marked in bold and italic)

Methods	1%	5%	10%	25%	30%	50%
iPiano-Net [24]	21.74/0.4864	25.21/0.6461	27.66/0.7670	31.62/0.9132	31.98/0.9004	36.64/0.9687
BCSnet [25]	21.94/0.5031	25.12/0.6637	*27.88*/0.7773	32.09/0.9182	32.76/0.9249	36.83/0.9731
AMP-Net [26]	21.82/0.5503	24.01/0.6671	26.41/0.7941	32.05/0.9195	33.24/0.9367	37.51/0.9750
COAST [27]	20.30/0.4073	23.18/0.6343	26.26/0.7682	30.49/0.8901	31.49/0.9096	35.32/0.9585
TransCS [28]	21.79/0.5045	25.03/0.7076	27.80/0.8282	31.99/*0.9257*	32.81/*0.9388*	37.31/0.9758
SODAS-Net [29]	21.92/*0.5584*	24.71/0.6810	26.41/0.7740	30.87/0.8984	32.03/0.9191	36.12/0.9644
PIPO-Net [30]	21.77/0.5304	25.04/0.7101	26.84/0.8003	32.11/0.9182	32.31/0.9152	37.56/*0.9772*
SuperTA-Net	*22.11*/0.5429	*25.22/0.7196*	27.74/*0.8289*	*32.23*/0.9192	*33.45*/0.9351	*37.84*/0.9711
SuperTA+RR-Net	**22.20/0.5677**	**25.50/0.7309**	**28.51/0.8455**	**32.77/0.9347**	**34.22/0.9451**	**38.32/0.9791**

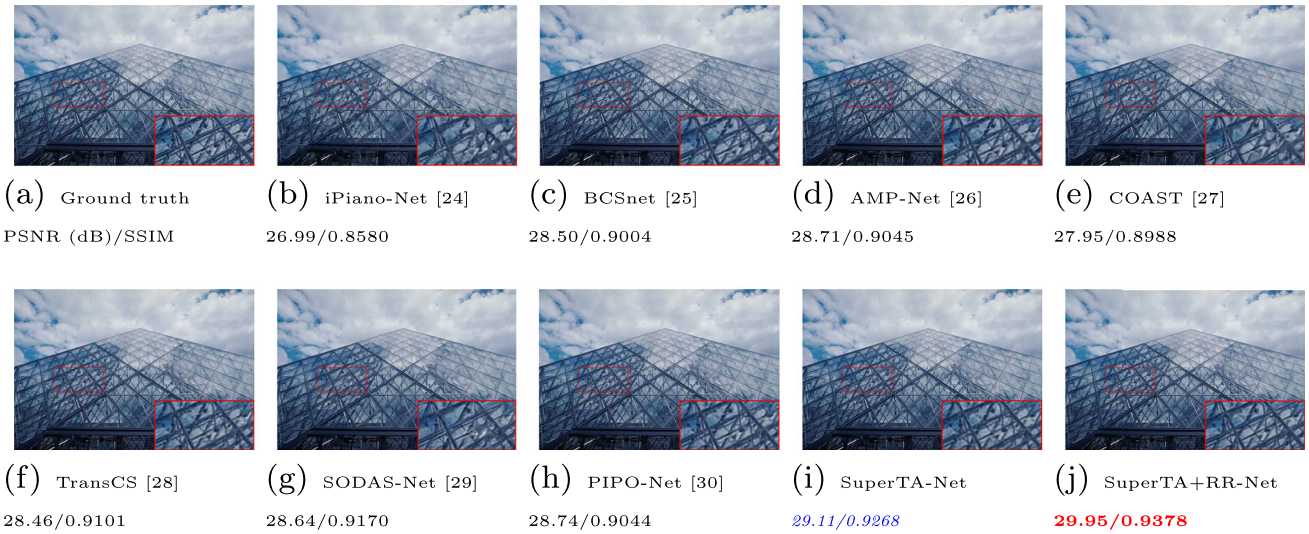

Fig. 6 Comparisons of reconstructed building images by nine methods with 30% CS ratio

Table 6 PSNR (dB)/SSIM results of the nine DUN methods on Set11 (the top and the second-best results are marked in bold and italic)

Methods	1%	5%	10%	25%	30%	50%
iPiano-Net [24]	19.16/0.5162	24.02/0.7534	28.12/0.8305	33.55/0.9472	34.82/0.9441	38.88/0.9702
BCSnet [25]	20.81/0.5427	26.31/0.7686	29.38/0.8637	34.20/0.9408	35.63/0.9495	39.57/0.9720
AMP-Net [26]	20.21/0.5581	25.33/0.7692	29.13/0.8610	34.64/0.9488	36.12/0.9602	40.32/0.9804
COAST [27]	20.32/0.5373	25.33/*0.7846*	28.74/0.8619	33.98/0.9407	35.11/0.9505	38.94/0.9744
TransCS [28]	*22.81*/0.5426	26.25/0.7728	29.54/0.8877	35.06/*0.9548*	35.62/0.9588	40.49/0.9815
SODAS-Net [29]	20.47/0.5323	26.24/0.7700	28.77/0.8612	34.23/0.9455	35.34/0.9523	39.88/0.9772
PIPO-Net [30]	21.43/0.5631	26.94/0.7773	29.77/0.9058	34.68/0.9500	36.32/0.9598	40.43/**0.9831**
SuperTA-Net	22.29/*0.5872*	*27.14*/0.7800	*30.37/0.9084*	*35.27*/0.9519	*36.62/0.9635*	*40.52*/0.9810
SuperTA+RR-Net	**23.73/0.6052**	**27.51/0.8157**	**30.53/0.9113**	**35.61/0.9594**	**36.92/0.9662**	**40.94**/*0.9827*

The last group of ablation studies is carried out to assess the effectiveness of the reinforcement network. In order to certify the applicability of the proposed RR-Net, three more CS reconstruction networks, ReconNet [16], ISTA-Net [22] and AMP-Net [26], are included for testing. For three CS ratios 1%, 10% and 25%, the results are shown in Table 4. It can be seen that the RR-Net is also suitable for reinforcing other networks besides the SuperTA-Net. Roughly speaking, for low CS ratio, e.g., 1%, if a network without the RR-Net integrated reconstructs the image better, more improvement could be achieved when the RR-Net is integrated. This may be explained by the fact that the RR-Net is lightweight and it is suitable for recovering relatively sparse residual image. For high CS ratio, e.g., 25%, even though the main network performs pretty well and the room for improvement is limited, the RR-Net can still bring gain.

Comparison experiments

Several groups of experiments based on datasets BSDS500 testing set, Set11 and BSD68 are carried out to test the performance of the proposed networks SuperTA-Net and SuperTA+RR-Net. The comparisons include seven recent DUN-based methods that are iPiano-Net [24], BCSnet [25], AMP-Net [26], COAST [27], TransCS [28], SODAS-Net [29] and PIPO-Net [30].

Table 5 summarizes the PSNR (dB)/SSIM results achieved by the nine methods with different CS ratios ranging from 1 to 50% on the BSDS500 testing set which contains 200 colorful images. The top and the second-best results in the table are marked in red bold and blue italic, respectively. It can be seen that the proposed SuperTA+RR-Net outperforms other methods in terms of PSNR and SSIM for all scenarios, and the main network SuperTA-Net achieves the second-best results in most cases. Taking the prominent AMP-Net [26] as the benchmark, the SuperTA-Net gets the PSNR gains of 0.29 dB, 1.21 dB, 1.33 dB, 0.18 dB, 0.21 dB and 0.33 dB over

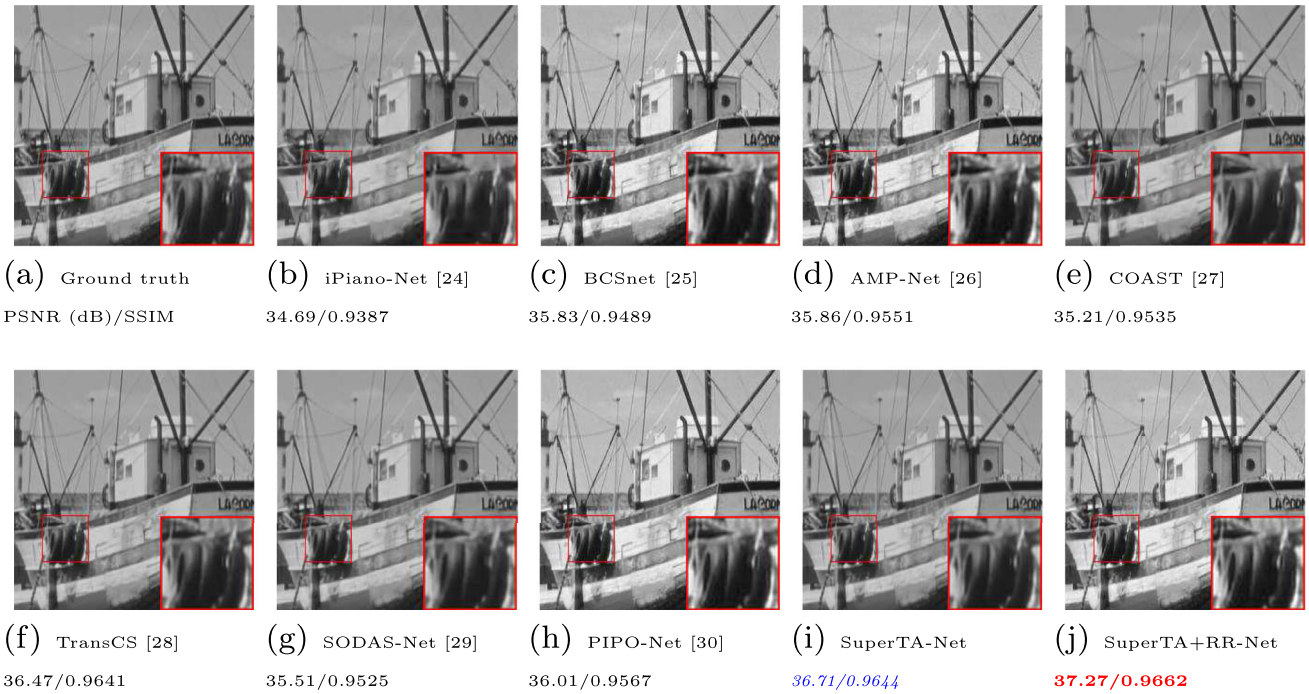

Fig. 7 Comparisons of reconstructed boat images by nine methods with 25% CS ratio

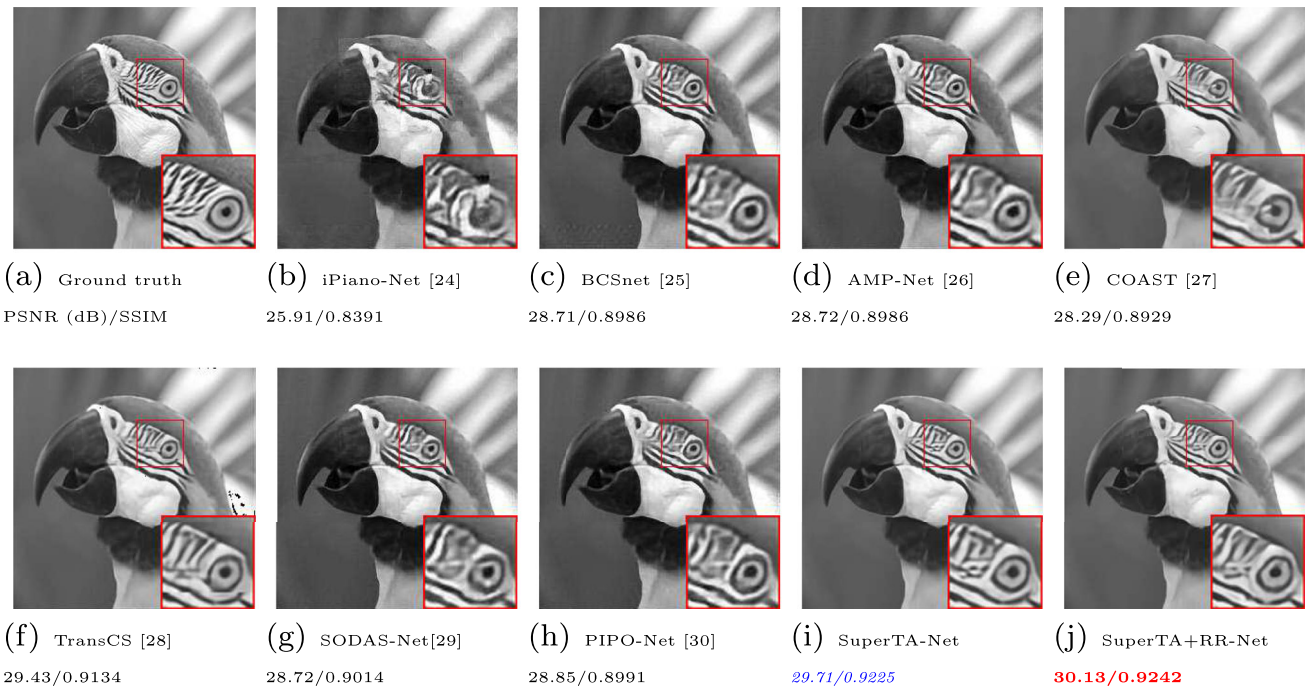

Fig. 8 Comparisons of reconstructed parrot images by nine methods with 10% CS ratio

the AMP-Net for CS ratio 1%, 5%, 10%, 25%, 30% and 50%, respectively, while the gains obtained by the SuperTA+RR-Net are 0.38 dB, 1.49 dB, 2.10 dB, 0.72 dB, 0.98 dB and 0.81 dB.

Take the building image which contains rich texture features as an example. The visual effects together with the PSNR (dB)/SSIM results of the reconstructed building images with 30% CS ratio are shown in Fig. 6. As can be observed from the amplifying parts, the proposed networks recover more details of the texture and the results seem more authentic, especially that of the SuperTA+RR-Net.

Table 7 PSNR (dB)/SSIM results of the nine DUN methods on BSD68 (the top and the second-best results are marked in bold and italic)

Methods	1%	5%	10%	25%	30%	50%
iPiano-Net [24]	20.67/0.5117	23.45/0.6472	28.34/0.7442	30.52/0.8831	31.33/0.8982	34.94/0.9519
BCSnet [25]	20.31/0.5319	25.23/0.6845	28.98/0.8715	31.29/0.9046	32.76/0.9341	36.69/0.9592
AMP-Net [26]	21.82/0.5571	24.81/0.6642	29.04/0.8525	32.01/0.9130	33.15/0.9352	36.59/0.9620
COAST [27]	20.13/0.5283	23.14/0.6208	26.28/0.7422	30.07/0.9041	31.06/0.8934	34.76/0.9487
TransCS [28]	21.76/0.5669	24.33/0.6639	29.88/0.8763	32.35/0.9146	33.45/0.9235	37.03/0.9624
SODAS-Net [29]	21.53/0.5444	24.65/0.6512	26.45/0.7428	30.51/0.8793	31.25/0.9011	35.61/0.9492
PIPO-Net [30]	22.09/0.5604	24.33/0.6721	*30.45*/0.9002	32.32/*0.9307*	33.30/*0.9384*	*37.20*/**0.9801**
SuperTA-Net	*22.50/0.5732*	*25.40*/**0.7074**	30.31/*0.9014*	*32.46*/0.9281	*33.48*/0.9372	37.19/0.9752
SuperTA+RR-Net	**22.56/0.5833**	**25.64**/*0.7031*	**30.91/0.9023**	**32.55/0.9314**	**33.62/0.9412**	**37.33**/*0.9779*

When the dataset Set11 consisting of 11 famous gray images is taken for experiments, the PSNR (dB)/SSIM results achieved by the nine networks with different CS ratios are summarized in Table 6. The results reveal similar conclusions and verify the superiority of the proposed SuperTA-Net and SuperTA+RR-Net.

Figures 7 and 8 show the visual effects together with the PSNR (dB)/SSIM results of reconstructed boat images with 25% CS ratio and reconstructed parrot images with 10% CS ratio, respectively. These results highlight the superior performance of the proposed networks.

Finally, the dataset BSD68 which contains more high-resolution images is chosen for certifying the performance. Although there exist some fluctuations, the performance of SuperTA-Net and SuperTA+RR-Net is generally satisfactory. This can be concluded from the results in Table 7.

Conclusion

In this paper, a novel deep unfolding network named SuperTA-Net has been proposed for image CS reconstruction. Under the similar framework of our previous work PIPO-Net, the multi-channel transmission strategy is adopted in each deep reconstruction module to augment the performance of the network by reducing the loss of critical information. Meanwhile, in order to avoid the issues such as high information redundancy and high computational burden raised by the multi-channel strategy, the attention based supervisor has been designed to evaluate the importance of each channel. The SuperTA-Net is sufficient for obtaining satisfactory image reconstruction result. In addition, a supplementary reinforcement network can be added to recover the residual between the original image and the output of SuperTA-Net. As the main component RR-Net is lightweight, the reinforcement network can be integrated to all kinds of CS reconstruction networks. Extensive experiments on image CS reconstruction have demonstrated the superior performance of the proposed networks.

Though the proposed networks learn the prior from data, CNN-based structures usually have limited receptive field. Hence, the transformer-based structures can be adopted to learn the prior which may consider longer-range dependencies among image patches. We leave this topic for the future work.

Author contribution statement Zhijie Zhang: Conceptualization, Simulation, Validation, Writing-Original draft Huang Bai: Conceptualization, Simulation, Validation, Writing-Original draft Ljubisa Stankovic: Investigation, reviewing Junmei Sun: Investigation, reviewing Xiumei Li: Methodology, Writing-Reviewing and Editing

Data Availability The data that support the findings of this study are available from the corresponding author upon reasonable request.

Declarations

Conflict of interest The authors declare that they have no Conflict of interest.

References

1. Bai H, Li G, Li S, Li Q, Jiang Q, Chang L (2015) Alternating optimization of sensing matrix and sparsifying dictionary for

compressed sensing. IEEE Trans Signal Process 63(6):1581–1594. https://doi.org/10.1109/TSP.2015.2399864
2. Duarte MF, Davenport MA, Takhar D, Laska JN, Sun T, Kelly KF, Baraniuk RG (2008) Single-pixel imaging via compressive sampling. IEEE Signal Process Mag 25(2):83–91. https://doi.org/10.1109/MSP.2007.914730
3. Liang D, Cheng J, Ke Z, Ying L (2020) Deep magnetic resonance image reconstruction: inverse problems meet neural networks. IEEE Signal Process Mag 37(1):141–151. https://doi.org/10.1109/MSP.2019.2950557
4. Jiang Q, Li S, Zhu Z, Bai H, He X, de Lamare RC (2020) Design of compressed sensing system with probability-based prior information. IEEE Trans Multimed 22(3):594–609. https://doi.org/10.1109/TMM.2019.2931400
5. Xu J, Yang J, Kimishima F, Taniguchi I, Zhou J (2023) Compressive sensing based image codec with partial pre-calculation. IEEE Trans Multimed. https://doi.org/10.1109/TMM.2023.3327534
6. Bai H, Li S, He X (2016) Sensing matrix optimization based on equiangular tight frames with consideration of sparse representation error. IEEE Trans Multimed 18(10):2040–2053. https://doi.org/10.1109/TMM.2016.2595261
7. Mallat SG, Zhang Z (1993) Matching pursuits with time-frequency dictionaries. IEEE Trans Signal Process 41(12):3397–3415. https://doi.org/10.1109/78.258082
8. Tropp JA, Gilbert AC (2007) Signal recovery from random measurements via orthogonal matching pursuit. IEEE Trans Inf Theory 53(12):4655–4666. https://doi.org/10.1109/TIT.2007.909108
9. Li C, Yin W, Jiang H, Zhang Y (2013) An efficient augmented Lagrangian method with applications to total variation minimization. Comput Optim Appl 56(3):507–530. https://doi.org/10.1007/s10589-013-9576-1
10. Blumensath T, Davies ME (2008) Iterative thresholding for sparse approximations. J Fourier Anal Appl 14(5):629–654. https://doi.org/10.1007/s00041-008-9035-z
11. Daubechies I, Defrise M, De Mol C (2004) An iterative thresholding algorithm for linear inverse problems with a sparsity constraint. Commun Pure Appl Math 57(11):1413–1457. https://doi.org/10.1002/cpa.20042
12. Beck A, Teboulle M (2009) A fast iterative shrinkage-thresholding algorithm for linear inverse problems. SIAM J Imaging Sci 2(1):183–202. https://doi.org/10.1137/080716542
13. Metzler CA, Maleki A, Baraniuk RG (2016) From denoising to compressed sensing. IEEE Trans Inf Theory 62(9):5117–5144. https://doi.org/10.1109/TIT.2016.2556683
14. Stanković L, Sejdić E, Stanković S, Daković M, Orović I (2019) A tutorial on sparse signal reconstruction and its applications in signal processing. Circuits Syst Signal Process 38:1206–1263. https://doi.org/10.1007/s00034-018-0909-2
15. Mousavi A, Patel AB, Baraniuk RG (2015) A deep learning approach to structured signal recovery. In: 2015 53rd Annu. Allerton Conf. Commun. Control Comput. (Allerton), Monticello, pp 1336–1343. https://doi.org/10.1109/ALLERTON.2015.7447163
16. Kulkarni K, Lohit S, Turaga P, Kerviche R, Ashok A (2016) ReconNet: non-iterative reconstruction of images from compressively sensed measurements. In: 2016 IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Las Vegas, pp 449–458. https://doi.org/10.1109/CVPR.2016.55
17. Yao H, Dai F, Zhang S, Zhang Y, Tian Q, Xu C (2019) DR^2-Net: deep residual reconstruction network for image compressive sensing. Neurocomputing 359:483–493. https://doi.org/10.1016/j.neucom.2019.05.006
18. Shi W, Jiang F, Liu S, Zhao D (2020) Image compressed sensing using convolutional neural network. IEEE Trans Image Process 29:375–388. https://doi.org/10.1109/TIP.2019.2928136
19. Monga KV, Li Y, Eldar YC (2021) Algorithm unrolling: interpretable, efficient deep learning for signal and image processing. IEEE Signal Process Mag 38(2):18–44. https://doi.org/10.1109/MSP.2020.3016905
20. Zhang J, Chen B, Xiong R, Zhang Y (2023) Physics-inspired compressive sensing: beyond deep unrolling. IEEE Signal Process Mag 40(1):58–72. https://doi.org/10.1109/MSP.2022.3208394
21. Gregor K, LeCun Y (2010) Learning fast approximations of sparse coding. In: 27th Int. Conf. Mach. Learn. (ICML), Madison, pp 399–406. https://doi.org/10.5555/3104322.3104374
22. Zhang J, Ghanem B (2018) ISTA-Net: interpretable optimization-inspired deep network for image compressive sensing. In: 2018 IEEE Conf. Comput. Vis. Pattern Recog. (CVPR), Salt Lake City, pp 1828–1837. https://doi.org/10.1109/CVPR.2018.00196
23. Yang Y, Sun J, Li H, Xu Z (2020) ADMM-CSNet: a deep learning approach for image compressive sensing. IEEE Trans Pattern Anal Mach Intell 42(3):521–538. https://doi.org/10.1109/TPAMI.2018.2883941
24. Su Y, Lian Q (2020) iPiano-Net: nonconvex optimization inspired multi-scale reconstruction network for compressed sensing. Signal Process Image Commun. https://doi.org/10.1016/j.image.2020.115989
25. Zhou S, He Y, Liu Y, Li C, Zhang J (2021) Multi-channel deep networks for block-based image compressive sensing. IEEE Trans Multimed 23:2627–2640. https://doi.org/10.1109/TMM.2020.3014561
26. Zhang Z, Liu Y, Liu J, Wen F, Zhu C (2021) AMP-Net: denoising-based deep unfolding for compressive image sensing. IEEE Trans Image Process 30:1487–1500. https://doi.org/10.1109/TIP.2020.3044472
27. You D, Zhang J, Xie J, Chen B, Ma S (2021) COAST: controllable arbitrary-sampling network for compressive sensing. IEEE Trans Image Process 30:6066–6080. https://doi.org/10.1109/TIP.2021.3091834
28. Shen M, Gan H, Ning C, Hua Y, Zhang T (2022) TransCS: a transformer-based hybrid architecture for image compressed sensing. IEEE Trans Image Process 31:6991–7005. https://doi.org/10.1109/TIP.2022.3217365
29. Song J, Zhang J (2023) SODAS-Net: side-information-aided deep adaptive shrinkage network for compressive sensing. IEEE Trans Instrum Meas. https://doi.org/10.1109/TIM.2023.3304676
30. Li X, Zhang Z, Bai H, Stanković L, Hao J, Sun J (2023) PIPO-Net: a penalty-based independent parameters optimization deep unfolding network. https://doi.org/10.48550/arXiv.2311.02443. arXiv:2311.02443
31. Li S, Li Q, Zhu Z, Tang G, Wakin MB (2020) The global geometry of centralized and distributed low-rank matrix recovery without regularization. IEEE Signal Process Lett 27:1400–1404. https://doi.org/10.1109/LSP.2020.3008876
32. Boyd S, Parikh N, Chu E, Peleato B, Eckstein J (2011) Distributed optimization and statistical learning via the alternating direction method of multipliers. Found Trends Mach Learn 3(1):1–122. https://doi.org/10.1561/2200000016
33. Ochs P, Chen Y, Brox T, Pock T (2014) iPiano: inertial proximal algorithm for nonconvex optimization. SIAM J Imaging Sci 7(2):1388–1419. https://doi.org/10.1137/130942954
34. Fowler J, Mun S, Tramel EW (2010) Block-based compressed sensing of images and video. Found Trends Signal Process 4(4):297–416. https://doi.org/10.1561/2000000033
35. Donoho DL, Maleki A, Montanari A (2009) Message-passing algorithms for compressed sensing. Proc Natl Acad Sci USA 106(45):18914–18919. https://doi.org/10.1073/pnas.0909892106
36. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser L, Polosukhin I (2017) Attention is all you need. In: 31st Conf. Neural Inf. Process. Syst. (NIPS), Long Beach
37. Nesterov Y (2013) Introductory lectures on convex optimization: a basic course. Springer, New York. https://doi.org/10.1007/978-1-4419-8853-9

38. Arbeláez P, Maire M, Fowlkes C, Malik J (2011) Contour detection and hierarchical image segmentation. IEEE Trans Pattern Anal Mach Intell 33(5):898–916. https://doi.org/10.1109/TPAMI.2010.161
39. Kingma DP, Ba J (2014) Adam: a method for stochastic optimization. https://doi.org/10.48550/arXiv.1412.6980. arXiv:1412.6980
40. Martin D, Fowlkes C, Tal D, Malik J (2001) A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In: 2001 IEEE Int. Conf. Comput. Vis. (ICCV), Vancouver, vol 2, pp 416–423. https://doi.org/10.1109/ICCV.2001.937655
41. Wang Z, Bovik AC, Sheikh HR, Simoncelli EP (2004) Image quality assessment: from error visibility to structural similarity. IEEE Trans Image Process 13(4):600–612. https://doi.org/10.1109/TIP.2003.819861