Single-Channel Signal Separation and Deconvolution with Generative Adversarial Networks

Qiuqiang Kong[1] , **Yong Xu**[2] , **Wenwu Wang**[1] , **Philip J.B. Jackson**[1] , **Mark D. Plumbley**[1]

[1]University of Surrey, Guildford, UK
[2]Tencent AI lab, Bellevue, USA

{q.kong, w.wang, p.jackson, m.plumbley}@surrey.ac.uk
lucayongxu@tencent.com

Abstract

Single-channel signal separation and deconvolution aims to separate and deconvolve individual sources from a single-channel mixture and is a challenging problem in which no prior knowledge of the mixing filters is available. Both individual sources and mixing filters need to be estimated. In addition, a mixture may contain non-stationary noise which is unseen in the training set. We propose a synthesizing-decomposition (S-D) approach to solve the single-channel separation and deconvolution problem. In synthesizing, a generative model for sources is built using a generative adversarial network (GAN). In decomposition, both mixing filters and sources are optimized to minimize the reconstruction error of the mixture. The proposed S-D approach achieves a peak-to-noise-ratio (PSNR) of 18.9 dB and 15.4 dB in image inpainting and completion, outperforming a baseline convolutional neural network PSNR of 15.3 dB and 12.2 dB, respectively and achieves a PSNR of 13.2 dB in source separation together with deconvolution, outperforming a convolutive non-negative matrix factorization (NMF) baseline of 10.1 dB.

1 Introduction

Single-Channel signal separation and deconvolution aims to separate and deconvolve sources from a single-channel mixture. One challenging aspect of single-channel signal separation and deconvolution is that only a single-channel mixture is available, so this problem is underdetermined. Second, there is no prior knowledge of the mixing filters of the sources. Both individual sources and mixing filters are unknown and need to be estimated. Third, there is no prior knowledge on the noise, which can be non-stationary and has not been seen in the training data. These difficulties lead to single-channel signal separation and deconvolution being a very challenging problem. Single-channel signal separation and deconvolution has many applications in image, speech and audio denoising [Xie *et al.*, 2012], inpainting [Yeh *et al.*, 2016], deconvolution and separation [Cichocki *et al.*, 2009; Mijovic *et al.*, 2010]. For example, an audio sensor usually receives signals from multiple sources convolved with channel distortion.

Much previous work focuses on source separation [Cichocki *et al.*, 2009; Grais *et al.*, 2014] or deconvolution [Levin *et al.*, 2009; Campisi and Egiazarian, 2017] independently, but not together. We categorize previous source separation and deconvolution methods into decomposition based approaches and regression based approaches. Decomposition methods usually learn a set of bases for sources and use these bases to decompose a mixture. Decomposing methods including non-negative matrix factorization (NMF) [Lee and Seung, 1999; Cichocki *et al.*, 2009; Kitamura *et al.*, 2013] assumes that a source can be represented by linear combination of a set of bases. NMF has been used in source representation and separation [Cichocki *et al.*, 2006; Kitamura *et al.*, 2013]. In contrast to the decomposition based approaches, regression based approaches learn a mapping from a mixture to an individual source. Such mappings can be modeled by neural networks. For example, Fully connected neural networks [Grais *et al.*, 2014] and convolutional neural networks (CNNs) [Jain and Seung, 2009; Zhang *et al.*, 2017]. In [Xie *et al.*, 2012], a stacked denoising auto-encoder (DAE) is proposed to recover sources from a mixture. CNNs are used for source deconvolution in [Xu *et al.*, 2014].

However, many decomposition methods such as NMF and ICA are shallow layer models, which are typically a linear combination of bases. These shallow layer models do not have enough capacity to represent a broad range of sources compared with neural networks [Jain and Seung, 2009]. On the other hand, regression based approaches such as deep neural networks are able to model complicated mappings but require both mixture and target sources for training. Regression based methods may not generalize well if the mixing filter and noise in the testing data have different distribution from the training data, which will result in poor separation results when the mixing filter and noise are unseen in the training data [Yosinski *et al.*, 2014]. Recently generative adversarial networks (GANs) have been proposed for solving the source separation problem [Fan *et al.*, 2018; Subakan and Smaragdis, 2017; Stoller *et al.*, 2017]. So far these methods assume that the mixing filters in the single-channel signal separation problem are known.

This paper proposed a novel synthesizing-decomposition

(S-D) approach to solve the single-channel source separation and deconvolution problem. Compared to the conventional regression approaches, the S-D approach applies generative adversarial network (GANs) to solve this problem in a generative way. The S-D approach can estimate both the sources and convolutive mixing filters, while conventional regression methods do not estimate convolutive mixing filters. In addition, we formulate the single-channel signal separation and deconvolution problem as a Bayesian maximum a posteriori (MAP) estimation which is a constrained non-convex optimization problem. In the S-D approach, a generative model is built for sources using a generative adversarial network (GAN). In decomposition, both sources and mixing filters can be obtained by minimizing the reconstruction error of a mixture. To tackle the non-convex optimization problem, repeating the decomposition with different initializations can significantly increase the underdetermined single-channel signal separation and deconvolution performance. We carry out the underdetermined single-channel signal separation and deconvolution experiments on MNIST dataset as a starting research to show the effectiveness of the proposed S-D approach with GANs .

This paper is organized as follows: Section 2 formulates the underdetermined single-channel signal separation and deconvolution problem. Section 3 proposes the synthesising-decomposition (S-D) approach for this problem. Section 4 shows experimental results. Section 5 concludes and forecasts future work.

2 Single-channel signal separation and deconvolution

In underdetermined single-channel signal separation and deconvolution, a single-channel mixture $x(u) \in L^2(\Omega), u \in \Omega$ is composed of individual sources $s_k(u) \in L^2(\Omega), u \in \Omega, k = 1, ..., K$ convolved with unknown filters $\alpha_k(u) \in L^2(\Omega), u \in \Omega, k = 1, ..., K$ followed by unknown additional noise $n(u) \in L^2(\Omega), u \in \Omega$. The space Ω can be a Euclidean space $\mathbb{R}^d$ where K and d denote the number and the dimension of sources, respectively:

$$x(u) = \sum_{k=1}^{K} (\alpha_k * s_k)(u) + n(u). \quad (1)$$

The symbol $*$ represents the convolution operation:

$$(\alpha_k * s_k)(u) = \int_{\mathbb{R}^d} \alpha_k(u-v)s_k(v)dv. \quad (2)$$

For the simple case of source separation without deconvolution, in 2) $\alpha_k(u)$ simplifies to $\alpha_k(u-v) = \alpha_k\delta(u-v)$ where $\delta(u)$ is the Dirac delta function. General single-channel signal separation and deconvolution problem concerns both separating and deconvolving individual sources $s_k(u), k = 1, ..., K$ from a single-channel mixture $x(u)$ while the mixing filters $\alpha_k(u), k = 1, ..., K$ and the noise signal $n(u)$ are unknown in (1). In the following paper, we simplify the notation of $x(u), s_k(u), \alpha_k(u)$ to x, s_k and α_k, respectively.

In the regression based approaches [Jain and Seung, 2009; Grais *et al.*, 2014], a mapping from a mixture to a source signal is modeled by deep neural networks and learned to separate the k-th source: $f_k : x \mapsto s_k$. In separation, separated

sources are obtained by forwarding a mixture to the model: $\hat{s}_{\text{test}} = f_k(x_{\text{test}})$. However there are the following problems of the regression based approaches:

Problem 1. In regression based supervised learning, the training data x_{train} and testing data x_{test} should have the same distribution, otherwise the trained model will be biased [Yosinski *et al.*, 2014]. However, in single-channel signal separation and deconvolution, no prior knowledge of test noise n is available. The model trained with training noise may not generalize well to sources with unseen non-stationary noise.

Problem 2. In single-channel signal separation and deconvolution, both the sources s_k and mixing filters α_k are unknown and need to be estimated.

Problem 3. Previous regression and decomposition based approaches do not constrain the distribution of the separated sources $\hat{s}$ to be the same as the distribution of real sources $p_{\text{real}}(s)$. Ideally, the separated sources $\hat{s}$ should be regularized in the area where $p_{\text{real}}(\hat{s})$ has larger value.

Decomposition approaches such as NMF can be trained on individual sources instead of on a mixture so that Problem 1 can be mitigated. Recently, GANs [Fan *et al.*, 2018; Subakan and Smaragdis, 2017; Stoller *et al.*, 2017] have been applied to source separation to solve Problem 3 to constrain separated sources be laid in natural source space. However, those methods are based on the assumption that the mixing filters α_k are constants so that they are solving only separation but not deconvolution problem as shown in (1).

3 Proposed synthesising-decomposition (S-D) approach

3.1 Maximum a posteriori (MAP) estimation

In this section, we first formulate the single-channel signal separation and deconvolution problem in (1) as a Bayesian parameter estimation problem. We denote $\theta = \{s_1, ..., s_K, \alpha_1, ..., \alpha_K\}$ as the set of parameters to be estimated, including sources and mixing filters. The estimated $\hat{\theta}$ can be obtained by maximum a posteriori (MAP) estimation:

$$\begin{aligned}
\hat{\theta} &= \underset{\theta}{\operatorname{argmax}} \ p(\theta|x) \\
&= \underset{\theta}{\operatorname{argmax}} \ p(x|\theta)p(\theta).
\end{aligned} \quad (3)$$

The first term $p(x|\theta)$ in (3) is a likelihood function. The reconstructed signal can be written as $\hat{x} = \sum_{k=1}^{K} \alpha_k * s_k$. Assuming n is a Gaussian process, the likelihood of observed signal given estimated signal can be written as:

$$\begin{aligned}
p(x|\theta) &= p(x|\hat{x}) \\
&= \prod_{u \in \Omega} p(x(u)|\hat{x}(u)) \\
&= \prod_{u \in \Omega} \mathcal{N}(x(u) - \sum_{k=1}^{K} \alpha_k * s_k, \sigma_n)
\end{aligned} \quad (4)$$

where $\mathcal{N}(\cdot, \cdot)$ is the probability density of a Gaussian distribution. The second term $p(\theta)$ in (3) is the prior probability of

Algorithm 1 Training of a GAN [Goodfellow *et al.*, 2014].

1: Inputs: Real data $s_n, n = 1, ..., N$.
2: Outputs: Parameters of the discriminator θ_d and the generator θ_g of a GAN.
3: **for** number of iterations **do**
  - Sample minibatch of m noise samples $\{z^{(1)}, ..., z^{(m)}\}$ from a Gaussian distribution $N(0, 1)$.
  - Sample minibatch of m examples $\{s^{(1)}, ..., s^{(m)}\}$ from real data.
  - Update the discriminator by ascending its stochastic gradient:

$$\nabla \theta_d \frac{1}{m} \sum_{i=1}^{m} \left[\log D(s^{(i)}) + \log(1 - D(G(z^{(i)}))) \right]$$

  - Sample minibatch of m noise samples $\{z^{(1)}, ..., z^{(m)}\}$ from a Gaussian distribution $N(0, 1)$.
  - Update the generator by descending its stochastic gradient:

$$\nabla \theta_g \frac{1}{m} \sum_{i=1}^{m} \log(1 - D(G(z^{(i)})))$$

4: **end for**

Algorithm 2 Decomposition of a mixture source. Hyperparameters: K: Number of individual sources.

1: Inputs: A mixture source. Generator G trained using algorithm 1.
2: Outputs: Separated and deconvolved sources $s_k, k = 1, ..., K$ and mixing filters $\alpha_k, k = 1, ..., K$.
3: Sample K seeds $\{z_1, ..., z_K\}$ and K mixing filters $\{\alpha_1, ..., \alpha_K\}$ from a uniform distribution $U(-0.1, 0.1)$.
4: **for** number of iterations **do**
  - Calculate reconstructed signal $\hat{s} = \sum_{k=1}^{K} \alpha_k * G(z_k)$.
  - Calculate gradient ∇_θ from equation (11) where $\phi = \{z_1, ..., z_K, \alpha_1, ..., \alpha_K\}$.
  - Update $\phi = \{z_1, ..., z_K, \alpha_1, ..., \alpha_K\}$ using Adam optimizer [Kingma and Ba, 2015] with a learning rate of 0.01, beta1=0.9 and beta2=0.999.

5: **end for**

θ. Assuming the sources and filters are independent of each other, we can write $p(\theta)$ as:

$$p(\theta) = \prod_{k=1}^{K} p(\alpha_k) \prod_{k=1}^{K} p(s_k). \tag{5}$$

We assume $s_k, k = 1, ..., K$ to have a compact support $V \in \Omega$. Substituting equation (4) and (5) to equation (3) the estimation of sources and filters can be obtained by solving the following optimization problem:

$$\hat{s}_k, ..., \hat{s}_K, \hat{\alpha}_1, ..., \hat{\alpha}_K = \underset{\substack{s_1 \in V, ..., s_K \in V \\ \alpha_1, ..., \alpha_J}}{\text{argmax}}$$
$$\prod_{u \in \Omega} \mathcal{N}\left(x(u) - \sum_{k=1}^{K} \alpha_k * s_k, \sigma_n\right) \prod_{k=1}^{K} p(\alpha_k) \prod_{k=1}^{K} p(s_k) \tag{6}$$

3.2 Optimization with S-D approach

To optimize (6) is difficult because of the constraint of $s_k \in V$. The source prior $p(s_k)$ is unknown, so that V can not be written in a closed form. Our solution is to convert (6) to an unconstrained optimization problem. In the proposed S-D approach, we first build a generative model for x_k with a GAN [Goodfellow *et al.*, 2014; Subakan and Smaragdis, 2017]. A GAN consists of a generator G and a discriminator D. The generator G is a mapping from any distribution p_z such as a Gaussian distribution $N(0, \sigma \mathbf{I})$ to a real distribution of sources. We call p_z a *seed distribution* and sample $z \sim p_z$ as seeds. The generator G is trained to generate

samples to fool the discriminator D. The discriminator D is trained to discriminate fake sources from real sources. In other words, the generator G and the discriminator D play the following two-player minimax game with value function $V(G, D)$ [Goodfellow *et al.*, 2014]:

$$\min_G \max_D V(D, G) =$$
$$\mathbb{E}_{s \sim p_{\text{data}}(s)}[\log D(x)] + \mathbb{E}_{z \sim p_z(z)}[\log(1 - D(G(z)))] \tag{7}$$

where p_{data} is the real data probability density. The training of the GAN is shown in Algorithm 1. The generator G and discriminator D are trained iteratively. If both G and D have enough capacity, then the generated source distribution will converge to p_{data} [Subakan and Smaragdis, 2017]. Once GAN is successfully trained, there is $G(z) \in V$ for all z. To solve the optimization problem in (6), we substitute $s_k = G(z_k)$ and optimize over z_k instead of s_k so that the constraint $s \in V$ is eliminated. Now the variables to be optimized are z_k and the mixing filters α_k. In addition, GAN does not predict the probability density $p(s_k)$ of s_k so the optimization of equation (6) intractable. To solve this problem, we approximate $p(s_k)$ with:

$$p(s_k) = \begin{cases} 0, & s_k \notin V \\ 1/|V|, & s_k \in V. \end{cases} \tag{8}$$

Equation (8) assumes the probability density $p(s_k)$ outside V is zero. It is not required to know the value of $|V|$ as it is eliminated when optimizing (6):

$$\hat{z}_k, ..., \hat{z}_K, \hat{\alpha}_1, ..., \hat{\alpha}_K = \underset{\substack{z_1, ..., z_K \\ \alpha_1, ..., \alpha_J}}{\text{argmax}}$$
$$\prod_{u \in \Omega} \mathcal{N}\left(x(u) - \sum_{k=1}^{K} \alpha_k * s_k, \sigma_n\right) \prod_{k=1}^{K} p(\alpha_k). \tag{9}$$

We assume the coefficients in α_k to be Gaussian $\alpha_k \sim \mathcal{N}(0, \sigma_\alpha)$. Taking the logarithm of (9) the optimization can

be written as:

$$\hat{z}_k, ..., \hat{z}_K, \hat{\alpha}_1, ..., \hat{\alpha}_K = \underset{\substack{z_1, ..., z_K \\ \alpha_1, ..., \alpha_J}}{\operatorname{argmin}}$$

$$\left\| x - \sum_{k=1}^{K} \alpha_k * G(z_k) \right\|_2^2 + \beta \sum_{k=1}^{K} \|\alpha_k\|_2^2 \qquad (10)$$

where $\beta = \sigma_n / \sigma_\alpha$ is a regularization term for (10).

3.3 Optimization

To solve (10), we apply a gradient based iterative approach. We denote $\phi = \{z_1, ..., z_K, \alpha_1, ..., \alpha_K\}$ where z_k and α_k need to be optimized. First we randomly initialize ϕ, then the gradients of ϕ are calculated by:

$$\nabla_\phi = \frac{\partial}{\partial \phi} \left\{ \left\| x - \sum_{k=1}^{K} \alpha_k * G(z_k) \right\|_2^2 + \beta \sum_{k=1}^{K} \|\alpha_k\|_2^2 \right\}. \qquad (11)$$

The parameters ϕ are optimized using Algorithm 2. Because G is a non-linear mapping, so (10) is a non-convex function over ϕ. The gradient based methods might reach a local minimum depending on the initialization of seeds. To mitigate this problem we repeat Algorithm 2 for L times and choose the one with least reconstruction error.

4 Experiments

In this section, we apply the proposed S-D method to solve underdetermined image single-channel signal separation and deconvolution problem. We carry out experiments on MNIST 10-digit dataset [LeCun *et al.*, 1998] as a starting research for this challenging problem and show the effectiveness of the proposed S-D method. With different types of unknown mixing filters α_k and unknown interference noise n, the problem of (1) can be categorized as image denoising, inpainting, completion, deconvolution and separation, as shown in Table 1. The symbol '-' represents any type of noise. Previous works usually focus on one of these problems such as denoising [Jain and Seung, 2009], inpainting [Xie *et al.*, 2012], deconvolution [Xu *et al.*, 2014] or separation [Subakan and Smaragdis, 2017]. In this paper we solve these problem together with the proposed S-D method. The PyTorch implementation of this paper is released[1].

4.1 Model configuration

In the proposed S-D approach, we model the synthesising procedure with a conditional deep convolutive GAN (cDC-GAN) [Mirza and Osindero, 2014; Radford *et al.*, 2015], which can stabilize the training of a GAN and can generate high quality images as shown in [Radford *et al.*, 2015]. A cDCGAN can be trained conditioned on digits [Mirza and Osindero, 2014]. A cDCGAN consists of a generator G and a discriminator D. The input to G consists of a seed sampled from a Gaussian distribution $N(0, \sigma \mathbf{I})$ together with the

[1] https://github.com/qiuqiangkong/gan_separation_deconvolution

encoding of a digit as a condition. The Gaussian distribution has a dimension of 100 following [Radford *et al.*, 2015]. The generator G has 4 transpose convolutional layers with number of feature maps of 256, 256, 128 and 1, respectively. Following [Radford *et al.*, 2015], batch normalization [Ioffe and Szegedy, 2015] and ReLU non-linearity are applied after each transpose convolutional layer. The output of G is an image which has the same size as the images in the training data. The discriminator D takes a fake or a real image as input. The discriminator D consists of 4 convolutional layers, with a sigmoid output representing the probability that the input to D is from real data instead of generated data. Following [Radford *et al.*, 2015], we use the Adam [Kingma and Ba, 2015] optimizer with a learning rate of 0.0002, a β_1 of 0.5 and a β_2 of 0.999 to train the generator. In decomposition, we freeze the trained generator G. We approximate $p(x|\hat{x})$ with a Gaussian distribution which works well in our experiment. We set β to 0.001 to regularize the mixing filters α_k to be searched. The filters α_k and z_k are randomly initialized and optimized with Adam optimizer with a learning rate of 0.01, a β_1 of 0.9 and a β_2 of 0.999 (Algorithm 2).

For comparison with regression based approaches, we apply a CNN [Xie *et al.*, 2012] consists 4 layers with batch normalization [Ioffe and Szegedy, 2015] and ReLU non-linearity. The number of layers and parameters are set to be the same as the discriminator D in the cDCGAN. The CNN is trained to regress from individual source with noise $s + n$ to individual source s. For comparison with decomposition based approaches, we train a dictionary for each of the 10 digits using NMF [Cichocki *et al.*, 2009] with Euclidean distance. Each dictionary consists of 20 bases which performs well in our experiment. In decomposition, the trained dictionaries are concatenated to form a dictionary of 200 bases which is then used to decompose the mixtures.

4.2 Evaluation

Following [Xie *et al.*, 2012; Xu *et al.*, 2014; Jain and Seung, 2009], we use peak signal to noise ratio (PSNR) to evaluate single-channel signal separation and deconvolution quality. The higher PSNR indicates the better reconstruction quality. PSNR is defined as:

$$\text{PSNR} = 20 \log_{10} \left(\frac{\text{MAX}_I}{\sqrt{\text{MSE}}} \right) \qquad (12)$$

where MAX_I is the maximum value of a noise-free image. MSE represents mean squared error between two images I and J with size of $m \times n$:

$$\text{MSE}(I, K) = \frac{1}{mn} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} (I(i,j) - J(i,j))^2 \qquad (13)$$

4.3 Denoising, inpainting and completion

Denoising, inpaining and completion are special case of single-channel signal separation and deconvolution problem where α_k is an unknown constant and n is unknown noise such as Gaussian noise, non-stationary noise or corruption of an image. The first and second rows of Fig. 1 show the clean and noisy images. The third to the fifth rows show the

Table 1: Category of single-channel signal separation and deconvolution problem with different noise and mixing filters.

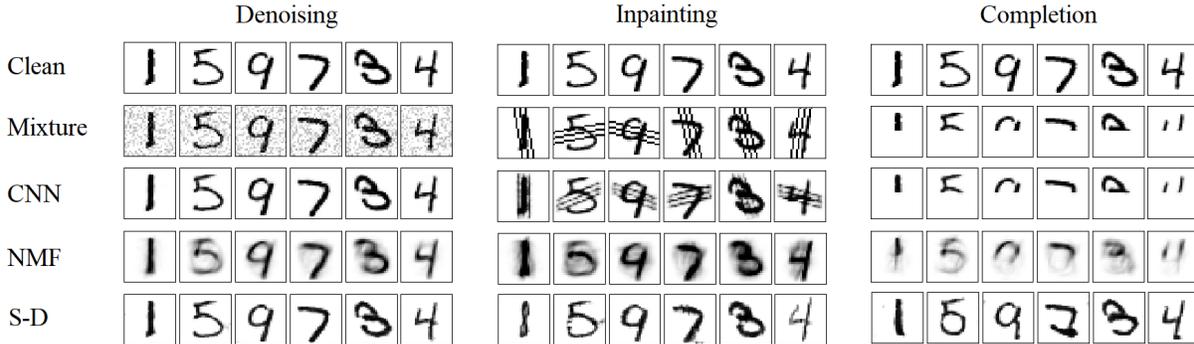|	Noise n	Mixing filters $\alpha_k, k = 1, ..., K$
Denoising	Gaussian	$K = 1$, α_k is a constant
Inpainting, Completion	Unknown	$K = 1$, α_k is a constant
Deconvolution	-	$K = 1$, α_k is a tensor
Separation	-	$K > 1$, α_k are constants
Separation + deconvolution	-	$K > 1$, α_k are tensors

Figure 1: Image denoising, inpainting and completion with CNN, NMF and S-D approaches.

Table 2: PSNR of image denoising, inpainting and completion with different approaches.

	denoising	inpainting	completion
CNN	**26.0 dB**	15.3 dB	12.2 dB
NMF	17.4 dB	13.4 dB	12.9 dB
convolutive NMF	18.3 dB	13.4 dB	13.0 dB
S-D with 1 init.	23.1 dB	15.2 dB	13.6 dB
S-D with 8 init.	25.1 dB	18.2 dB	**15.4 dB**
S-D with 32 init.	25.1 dB	**18.9 dB**	15.4 dB

Table 3: PSNR of image separation and deconvolution with different approaches.

	deconv.	sep.	sep. + deconv.
NMF	15.3 dB	9.4 dB	8.7 dB
convolutive NMF	18.3 dB	14.2 dB	10.1 dB
S-D with 1 init.	17.3 dB	13.7 dB	9.3 dB
S-D with 8 init.	21.9 dB	16.8 dB	11.5 dB
S-D with 32 init.	**23.2 dB**	**18.5 dB**	**13.2 dB**

denoised images with CNN, NMF and the proposed S-D approach. In the first column, testing noise and training noise have the same distribution so CNN performs well. However CNN based denoising methods do not generalize well to unseen noise such as non-stationary noise or image corruption shown in the second and third columns in Fig. 1. NMF performs better than CNN under unseen noise but sometimes produce unnatural separation result, which is Problem 3 we stated in Section 2. S-D approach has a good performance in all of image denoising, inpainting and completion. Table 2 shows PSNR of CNN, NMF, convolutive NMF and S-D approaches. S-D approach achieves a PSNR of 25.1 dB in image denoising which is comparable to CNN. NMF and convolutive NMF achieves similar PSNR of 17.4 dB and 18.3 dB, respectively. In image inpainting, S-D achieves a PSNR of 18.9 dB, outperforming NMF and CNN methods of 13.4 dB and 15.3 dB, respectively. This result shows source separation with S-D generalize well to unseen noise than NMF and CNN. In image completion, S-D approach achieves a PSNR of 15.4 dB, outperforming convolutive CNN of 12.2 dB and convolutive NMF of 12.9 dB respectively. Table 2 also shows

the decomposition in S-D approach with respect to the number of initializations. With 8 or 32 initializations the performance is 2 dB better than with only 1 initialization. This may result from the fact that the optimization problem in (10) is non-convex. Algorithm 2 is a gradient based method which may lead to the solution being in a local minimum. Repeating Algorithm 2 several times with different initializations and choosing the solution with least reconstruction error shows better performance.

4.4 Separation and deconvolution

We evaluate single-channel signal separation and deconvolution with the mixing filters $\alpha_k, k = 1, ..., K$ as unknown tensors, which is a very challenging task. In this case both of the mixing tensors α_k and individual sources s_k need to be estimated. Fig. 2 shows a mixture obtained by convolving clean sources with mixing filters followed by summation. In our experiment we set $K = 2$ and each mixing filter has a size of 5×5. In actual application scenarios the size of mixing filter depends on the task. Fig. 2 shows NMF based separation often lead to unnatural images. The S-D based approach can separate images with high quality and both the

Figure 2: Image separation and deconvolution with NMF and S-D approach.

sources s_k and mixing filters α_k can be estimated. Fig. 2 shows both estimated sources and mixing filters are learned correctly compared with the ground truth sources and mixing filters. The first column of Table 3 shows the results of image deconvolution without separation where K=1 and α is an unknown tensor. S-D achieves a PSNR of 23.2 dB and performs better than NMF and the convolutive NMF of 15.3 dB and 18.3 dB, respectively. The second column of Table 3 shows the results of image separation where α_k are unknown constants and $K = 2$. S-D achieves a PSNR of 18.5 dB and performs better than NMF and convolutive NMF of of 9.4 dB and 14.2 dB, respectively. The third column of Table 3 shows both of source separation and deconvolution where α_k are unknown tensors and $K = 2$. S-D achieves a PSNR of 13.2 dB and outperforms NMF and convolutive NMF of 8.7 dB and 10.1 dB, respectively. S-D with 32 initializations has higher PSNR than 8 initializations than 1 initialization, which shows the effectiveness of repeating Algorithm 2 several times to solve the non-convex optimization problem in (10).

5 Conclusion

In this paper, we propose a synthesis-decomposition (S-D) approach to solve single-channel signal separation and deconvolution problem. In synthesizing, a generative model for source signals is trained using a generative adversarial network (GAN). In decomposition, both sources and filters are optimized to minimize the reconstruction error. Instead of optimizing sources directly, we optimize over the seeds of a GAN. The proposed S-D approach achieves a PSNR of 18.9 dB and 15.4 dB in image inpainting and completion, outper-

forming the regression approach CNN and decomposition approach NMF. The S-D approach achieves a PSNR of 13.2 dB in image source separation with deconvolution, outperforming NMF of 8.7 dB. Repeating the decomposition in S-D several times can significantly improve PSNR. In future, we will explore the S-D approach to more source separation and deconvolution problems.

6 Acknowledgement

References

[Campisi and Egiazarian, 2017] Patrizio Campisi and Karen Egiazarian. *Blind image deconvolution: theory and applications*. CRC press, 2017.

[Cichocki *et al.*, 2006] A. Cichocki, R. Zdunek, and S. Amari. New algorithms for non-negative matrix factorization in applications to blind source separation. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2006.

[Cichocki *et al.*, 2009] A. Cichocki, R. Zdunek, A. H. Phan, and S. Amari. *Nonnegative matrix and tensor factorizations: applications to exploratory multi-way data analysis and blind source separation*. John Wiley & Sons, 2009.

[Fan *et al.*, 2018] Z. Fan, Y. Lai, and J. Jang. SVSGAN: Singing voice separation via generative adversarial net-

work. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2018.

[Goodfellow *et al.*, 2014] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio. Generative adversarial nets. In *Advances in Neural Information Processing Systems (NIPS)*, pages 2672–2680, 2014.

[Grais *et al.*, 2014] E. M. Grais, M. Sen, and H. Erdogan. Deep neural networks for single channel source separation. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 3734–3738, 2014.

[Ioffe and Szegedy, 2015] S. Ioffe and C. Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *International Conference on Machine Learning (ICML)*, 2015.

[Jain and Seung, 2009] V. Jain and S. Seung. Natural image denoising with convolutional networks. In *Advances in Neural Information Processing Systems (NIPS)*, pages 769–776, 2009.

[Kingma and Ba, 2015] D. P. Kingma and J. Ba. Adam: A method for stochastic optimization. In *International Conference on Learning Representations (ICLR)*, 2015.

[Kitamura *et al.*, 2013] D. Kitamura, H. Saruwatari, K. Shikano, K. Kondo, and Y. Takahashi. Music signal separation by supervised nonnegative matrix factorization with basis deformation. In *International Conference on Digital Signal Processing (DSP)*, 2013.

[LeCun *et al.*, 1998] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998.

[Lee and Seung, 1999] D. D. Lee and H. S. Seung. Learning the parts of objects by non-negative matrix factorization. *Nature*, 1999.

[Levin *et al.*, 2009] Anat Levin, Yair Weiss, Fredo Durand, and William T Freeman. Understanding and evaluating blind deconvolution algorithms. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2009.

[Mijovic *et al.*, 2010] B. Mijovic, M. De Vos, I. Gligorijevic, J. Taelman, and S. Van Huffel. Source separation from single-channel recordings by combining empirical-mode decomposition and independent component analysis. *IEEE Transactions on Biomedical Engineering*, 2010.

[Mirza and Osindero, 2014] Mehdi Mirza and Simon Osindero. Conditional generative adversarial nets. *arXiv preprint arXiv:1411.1784*, 2014.

[Radford *et al.*, 2015] A. Radford, L. Metz, and S. Chintala. Unsupervised representation learning with deep convolutional generative adversarial networks. *arXiv preprint arXiv:1511.06434*, 2015.

[Stoller *et al.*, 2017] D. Stoller, S. Ewert, and S. Dixon. Adversarial semi-supervised audio source separation applied to singing voice extraction. *arXiv preprint arXiv:1711.00048*, 2017.

[Subakan and Smaragdis, 2017] Cem Subakan and Paris Smaragdis. Generative adversarial source separation. *arXiv preprint arXiv:1710.10779*, 2017.

[Xie *et al.*, 2012] J. Xie, L. Xu, and E. Chen. Image denoising and inpainting with deep neural networks. In *Advances in Neural Information Processing Systems (NIPS)*, pages 341–349, 2012.

[Xu *et al.*, 2014] L. Xu, J. Ren, C. Liu, and J. Jia. Deep convolutional neural network for image deconvolution. In *Advances in Neural Information Processing Systems (NIPS)*, pages 1790–1798, 2014.

[Yeh *et al.*, 2016] R. Yeh, C. Chen, T. Y. Lim, M. Hasegawa-Johnson, and M. N. Do. Semantic image inpainting with perceptual and contextual losses. *arXiv preprint arXiv:1607.07539*, 2016.

[Yosinski *et al.*, 2014] J. Yosinski, J. Clune, Y. Bengio, and H. Lipson. How transferable are features in deep neural networks? In *Advances in Neural Information Processing Systems (NIPS)*, pages 3320–3328, 2014.

[Zhang *et al.*, 2017] K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang. Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising. *IEEE Transactions on Image Processing*, 26(7):3142–3155, 2017.