A Survey on Single-Channel Blind Source Separation in Communications

Mahan Nourifard[1]

[1]Electrical Engineering Department of Sharif, University of Technology

April 18, 2025

Abstract

Single-channel Blind Source Separation (SCBSS) is a crucial challenge in signal processing, aimed at recovering multiple source signals from a mixture observed through a single sensor, without prior knowledge of the mixing process. This problem is particularly relevant in various applications, including audio signal recovery, wireless communications, and speech enhancement. Traditional SCBSS methods, such as Independent Component Analysis (ICA) and Non-Negative Matrix Factorization (NMF), have demonstrated effectiveness but are often limited by computational complexity and the inability to handle more intricate, real-world signals. With the rise of deep learning, SCBSS has seen substantial advancements, as deep neural networks (DNNs) and other machine learning techniques have significantly enhanced the performance and robustness of separation algorithms. This survey provides a comprehensive review of the evolution of SCBSS, focusing on both classical methods and the more recent integration of artificial intelligence (AI) and deep learning approaches. Key advancements include the application of convolutional networks, recurrent neural networks (RNNs), and probabilistic binary mask techniques, which have been applied to speech separation and wireless communication signals. This Paper highlights the trade-offs between separation quality and computational efficiency, the challenges associated with signal mixtures, and the limitations of existing models. The review concludes by identifying gaps in current research, particularly the need for models that can handle more complex, non-linear mixtures in diverse environments. Future directions are proposed, focusing on improving the adaptability, noise robustness, and scalability of SCBSS systems, ensuring their applicability across a wide range of real-world scenarios.

A Survey on Single-Channel Blind Source Separation in Communications

Mahan Nourifard
Electrical Engineering Department of Sharif University of Technology
Mahan.Nourifard@ee.sharif.edu

***Abstract*—Single-channel Blind Source Separation (SCBSS) is a crucial challenge in signal processing, aimed at recovering multiple source signals from a mixture observed through a single sensor, without prior knowledge of the mixing process. This problem is particularly relevant in various applications, including audio signal recovery, wireless communications, and speech enhancement. Traditional SCBSS methods, such as Independent Component Analysis (ICA) and Non-Negative Matrix Factorization (NMF), have demonstrated effectiveness but are often limited by computational complexity and the inability to handle more intricate, real-world signals. With the rise of deep learning, SCBSS has seen substantial advancements, as deep neural networks (DNNs) and other machine learning techniques have significantly enhanced the performance and robustness of separation algorithms.**

This survey provides a comprehensive review of the evolution of SCBSS, focusing on both classical methods and the more recent integration of artificial intelligence (AI) and deep learning approaches. Key advancements include the application of convolutional networks, recurrent neural networks (RNNs), and probabilistic binary mask techniques, [4] which have been applied to speech separation and wireless communication signals. This Paper highlights the trade-offs between separation quality and computational efficiency, the challenges associated with signal mixtures, and the limitations of existing models.

The review concludes by identifying gaps in current research, particularly the need for models that can handle more complex, non-linear mixtures in diverse environments. Future directions are proposed, focusing on improving the adaptability, noise robustness, and scalability of SCBSS systems, ensuring their applicability across a wide range of real-world scenarios.

***Index Terms*—Single-channel Blind Source Separation, Deep Learning, Convolutional Neural Networks, Independent Component Analysis, Non-Negative Matrix Factorization, Speech Separation, Wireless Communications, Signal Processing, Probabilistic Binary Mask, Time-Domain Networks, Complex Domain Networks, Signal-to-Noise Ratio, Source Separation Algorithms, Deep Neural Networks, Computational Efficiency, Machine Learning, Cocktail Party Problem, Speech Enhancement, Signal Recovery, AI-based Separation Methods, Source Signal Estimation.**

I. INTRODUCTION

Single-Channel Blind Source Separation (SCBSS) is a critical problem in signal processing, where the aim is to recover source signals from a single mixed observation, with no prior knowledge about either the sources or the mixing process. [1] The observed mixture is the result of multiple source signals that are combined in an unknown way. This challenge is particularly difficult because, unlike multi-channel separation, where multiple sensor inputs provide information about the spatial characteristics of the sources, SCBSS relies on just one signal, often leading to underdetermined situations. The term "blind" refers to the lack of any information regarding the sources or how they were mixed, making SCBSS a significant challenge in signal processing.

SCBSS is a foundational problem in many areas of signal processing, with broad applications across a variety of domains such as audio processing, speech enhancement, communication systems, and medical signal processing. In everyday life, SCBSS techniques can be seen in action when we filter out background noise to focus on a single speaker in a noisy environment, such as a cocktail party or in a telecommunications setting. This scenario, often referred to as the "cocktail party problem," has been a benchmark for research in SCBSS. However, the same challenge persists in many practical applications, such as enhancing voice clarity in mobile communications or extracting specific sources from environmental noise for medical diagnostics.

The problem becomes even more complex when considering that sources in a mixture often overlap in both time and frequency domains, making the task of separation incredibly difficult. In SCBSS, the problem becomes one of "blind" separation, where there is no explicit knowledge of the mixing process or the underlying sources. This is different from "non-blind" source separation, where some prior knowledge or assumptions about the sources or mixing process is available. The absence of such knowledge makes SCBSS a significantly harder problem to solve, often requiring sophisticated algorithms to isolate and reconstruct the original signals.

Traditional approaches to SCBSS have relied heavily on statistical and signal processing methods, such as Independent Component Analysis (ICA). ICA assumes that the source signals are statistically independent and uses this assumption to recover the sources from the observed mixture. While ICA has been highly successful in certain conditions, particularly when the number of sources is equal to or fewer than the number of sensors, it has limitations in more complex scenarios. For instance, ICA struggles when sources overlap heavily in both time and frequency or when the mixing process is nonlinear. Moreover, in real-world applications, noise and other environmental factors complicate the separation process, requiring more advanced models to handle these challenges.

As the limitations of traditional methods became apparent, research began to explore more flexible and powerful techniques, particularly those based on artificial intelligence

(AI) and machine learning (ML). [1] The application of machine learning models, especially deep learning (DL), has transformed the field of SCBSS in recent years. Deep learning techniques, such as Convolutional Neural Networks (CNNs), Long Short-Term Memory (LSTM) networks, [7] and Recurrent Neural Networks (RNNs), have shown tremendous promise in learning complex features from data and have outperformed traditional methods in many tasks, including SCBSS. These models are capable of automatically learning to separate source signals by training on large datasets, making them highly adaptable to a variety of signal mixtures.

One of the most significant breakthroughs in SCBSS has been the use of time-frequency (T-F) masking techniques, [4] such as the ideal binary mask (IBM). [4] The IBM approach works by applying a binary mask to a mixture's time-frequency representation, assigning each time-frequency unit to the target source or to noise. The ideal binary mask represents the theoretical best separation that can be achieved by distinguishing between the source signals in the time-frequency domain. Although the IBM is idealized and difficult to achieve in practice, deep learning models trained on large datasets can approximate the IBM, leading to substantial improvements in source separation. Recent work has demonstrated that deep neural networks (DNNs), particularly convolutional networks, can approach the ideal binary mask performance in real-world scenarios, such as speech separation [5] in noisy environments.

Another important advancement has been the development of complex-valued deep learning networks, which are designed to work directly with complex signals, such as those encountered in wireless communication. In traditional SCBSS, most methods operate on real-valued signals and require transformation into time-frequency representations. However, in wireless communications and many other domains, [3] signals are naturally complex-valued, containing both real and imaginary components. Techniques such as the Complex Time-Domain Dilated Convolutional Recurrent Network (CTDCRN) have been introduced to handle these signals directly in their complex form. These methods are able to more effectively separate mixed communication signals, demonstrating superior performance over traditional methods and even over real-valued networks in certain scenarios.

Despite these advancements, several challenges remain in SCBSS, particularly in the context of real-world applications. One of the primary challenges is dealing with mixtures where the number of sources exceeds the number of available sensors, a situation often referred to as an underdetermined mixing problem. In these cases, traditional methods such as ICA struggle to find a unique solution, and new approaches are required. Additionally, the assumption of statistical independence of sources, which is central to many SCBSS techniques, is often violated in practical settings, making the separation process even more difficult.

Another significant issue is the computational complexity of deep learning-based SCBSS models. While deep learning has dramatically improved performance, these models can be computationally expensive, particularly when training on large datasets or operating in real-time systems. Techniques to reduce the complexity, such as low-complexity architectures or efficient training strategies, are an active area of research. Furthermore, the generalization of these models remains a challenge, as models that perform well on training data may struggle to generalize to unseen mixtures or environments, limiting their real-world applicability.

The future of SCBSS lies in further advancements in deep learning techniques, particularly in improving their robustness to noise, interference, and complex source mixtures. Hybrid models that combine traditional signal processing techniques with the flexibility and adaptability of deep learning could provide a powerful solution to many of the remaining challenges. Moreover, incorporating domain-specific knowledge, such as prior information about the source signals or the mixing process, could further enhance the performance of SCBSS systems. These advancements will likely play a crucial role in the continued development of SCBSS techniques and their applications in various domains, from speech recognition and telecommunications to medical imaging and beyond.

This survey provides a comprehensive review of the state-of-the-art methods in SCBSS, with a particular focus on artificial intelligence-based approaches. We examine the theoretical foundations of SCBSS, including classical methods like ICA, as well as the recent advances that have incorporated deep learning into the problem-solving framework. By reviewing these methods, we aim to highlight the progress made in the field, identify the key challenges that remain, and explore potential directions for future research in SCBSS.

II. Background and Problem Formulation

Single-Channel Blind Source Separation (SCBSS) is a fundamental problem in the field of signal processing. The task of SCBSS involves recovering a set of source signals from a mixture of these signals observed through a single sensor, without having any prior information about the mixing process or the sources themselves. This makes SCBSS distinct from other source separation problems where multiple sensors or prior knowledge of the sources are available.

Mathematically, the problem can be framed as follows: given a mixture signal $y(t)$, which is the result of the linear superposition of r source signals $s_1(t), s_2(t), \dots, s_r(t)$, the goal is to estimate the individual source signals $s_1(t), s_2(t), \dots, s_r(t)$ from the observed mixture signal $y(t)$. The mixture can be expressed as:

$$y(t) = s_1(t) + s_2(t) + \cdots + s_r(t), \tag{1}$$

where $y(t)$ is the observed mixture, and $s_1(t), s_2(t), \dots, s_r(t)$ are the unknown source signals. The challenge in SCBSS arises from the fact that the number of sources r is often unknown, and the mixing process is assumed to be unknown as well.

In a practical setting, the signals are often mixed in both time and frequency domains, which makes it difficult to distinguish between the sources. Moreover, since only one mixture signal is available, the problem becomes underdetermined when the number of sources exceeds the number of sensors.

This is commonly encountered in scenarios like speech separation, [5] where multiple speakers are simultaneously talking, or in the case of environmental signal mixtures, such as music and background noise.

SCBSS is called ”blind” because there is no explicit information about the source signals or the mixing process. The only available information is the observed mixture signal. This sets SCBSS apart from ”non-blind” source separation, where prior knowledge of the sources, such as their characteristics or the mixing process, is available.

The problem of SCBSS is of great interest because of its wide range of applications, including:

- **Speech enhancement and separation**: Isolating speech from background noise or multiple speakers.
- **Audio signal processing**: Separating individual audio components from a mixture.
- **Wireless communications**: Enhancing signal recovery by separating communication signals in noisy environments.
- **Medical signal processing**: Extracting signals like fetal heartbeats or brain waves from mixed data.

The SCBSS problem has been studied extensively, and numerous methods have been proposed to address it. These methods range from traditional statistical techniques, such as Independent Component Analysis (ICA), to more recent approaches based on machine learning and deep learning.

However, despite these advancements, several challenges remain in SCBSS:

- **Non-linearity in mixing processes**: Real-world mixing processes are often non-linear, making it difficult for traditional linear models like ICA to perform well.
- **Overlapping sources**: When sources overlap both in time and frequency, the problem becomes more complex, and simple models fail to separate them effectively.
- **Noise and interference**: External noise and interference can severely degrade the performance of source separation algorithms, particularly in low Signal-to-Noise Ratio (SNR) conditions.
- **Computational complexity**: Advanced algorithms, particularly those based on deep learning, often require significant computational resources, which can be a barrier in real-time applications.

In this section, we outline the key challenges in SCBSS and introduce the formulations that have driven the development of various separation techniques, from classical approaches to modern AI-based methods. We also highlight how these approaches attempt to address the core problem of separating sources when the mixing process is unknown and no prior information is available.

III. Traditional Techniques for SCBSS

Single-Channel Blind Source Separation (SCBSS) has been extensively studied over the years, with traditional techniques forming the foundation of this research. These methods rely on statistical assumptions and mathematical transformations to separate the source signals from the observed mixed signal. This section explores some of the most widely used traditional techniques for SCBSS, including *Independent Component Analysis* (ICA), *Non-Negative Matrix Factorization* (NMF), and other statistical methods.

A. Independent Component Analysis (ICA)

Independent Component Analysis (ICA) is one of the most widely used methods for blind source separation. The key assumption in ICA is that the source signals are statistically independent. Given a mixture of these independent signals, ICA aims to recover the original sources by finding a linear transformation that maximizes the statistical independence between the separated signals.

Let $y(t)$ be the observed mixed signal, which is a linear combination of independent source signals $s_1(t), s_2(t), \ldots, s_r(t)$. The general form of the observed signal can be expressed as:

$$y(t) = A \cdot s(t)$$

where $y(t)$ is the observed mixture vector, A is the mixing matrix, and $s(t)$ is the vector of source signals. The goal of ICA is to estimate the source signals $\hat{s}(t)$ by finding a separating matrix W such that:

$$\hat{s}(t) = W \cdot y(t)$$

where W is the inverse of the mixing matrix A, and $\hat{s}(t)$ represents the separated source signals. ICA achieves this by maximizing the statistical independence of the estimated sources, typically using measures like kurtosis or negentropy, which quantify non-Gaussianity.

However, ICA assumes that the sources are independent, and it struggles with highly correlated sources or when the number of sources exceeds the number of available sensors (underdetermined mixtures).

B. Non-Negative Matrix Factorization (NMF)

Non-Negative Matrix Factorization (NMF) is another popular technique for SCBSS, particularly in audio source separation. NMF works by decomposing a non-negative matrix, such as a spectrogram of the mixed signal, into two smaller non-negative matrices that represent the sources and their corresponding activations over time.

Let V be the matrix representing the mixed signal, where each element of V is non-negative. The NMF algorithm aims to factorize V into two non-negative matrices D (the dictionary or basis functions) and H (the activations or coefficients) such that:

$$V \approx D \cdot H$$

where D contains the basis elements, which are patterns that represent the sources in the frequency domain, and H represents how these basis elements combine over time to reconstruct the observed signal. This factorization is typically achieved by minimizing the reconstruction error using optimization techniques, with the most common objective functions being the Frobenius norm or Kullback-Leibler divergence:

$$\min_{D,H} \|V - D \cdot H\|_F^2$$

where $\|\cdot\|_F$ denotes the Frobenius norm. Sparsity constraints can also be added to H to improve source separation by promoting sparse representations of the sources:

$$\min_{D,H} \|V - D \cdot H\|_F^2 + \lambda \|H\|_1$$

where $\|H\|_1$ is the l_1-norm of H, promoting sparsity, and λ is a regularization parameter.

NMF is advantageous in SCBSS due to its ability to model the non-negativity of audio sources, such as speech or music. However, NMF assumes a linear mixing model and is sensitive to the quality of the learned dictionary D.

C. Sparse Component Analysis (SCA)

Sparse Component Analysis (SCA) assumes that the source signals are sparse, meaning that most of the signal values are zero or close to zero at any given time. This assumption is particularly useful in applications like speech separation, [5] where only a few speech sources are active at any given time, and the rest of the sources are silent or inactive.

SCA aims to find a sparse representation of the mixed signals by enforcing sparsity in the separation process. The objective is to recover the sparse components of the observed mixture by minimizing the following objective function:

$$\min_{W} \|y(t) - W \cdot s(t)\|_2^2 + \lambda \|s(t)\|_1$$

where $\|s(t)\|_1$ is the l_1-norm, enforcing sparsity on the source signals. The term $\|y(t) - W \cdot s(t)\|_2^2$ represents the reconstruction error, and λ is the sparsity regularization parameter.

By assuming that only a small subset of sources are active at any given time, SCA can effectively recover the active sources while discarding inactive ones. However, it relies heavily on the sparsity assumption, which may not hold for all types of signals.

D. Other Traditional Techniques

Several other traditional techniques have been proposed for SCBSS, such as:

- **Principal Component Analysis (PCA)**: PCA can be used for dimensionality reduction and source separation by projecting the observed mixture onto the principal components that capture the most variance. The principal components are computed by solving the following eigenvalue problem:

  $$\text{Cov}(y(t)) \cdot v = \lambda v$$

  where $\text{Cov}(y(t))$ is the covariance matrix of the mixed signal, v is the eigenvector, and λ is the eigenvalue. While PCA is useful in some cases, it assumes the sources are uncorrelated, which may not always hold.
- **Blind Source Separation using Bayesian Inference**: This technique models the separation problem probabilistically, where the observed signal is treated as a mixture of source signals, and the goal is to infer the most likely source signals using Bayesian methods. The model is typically trained using Maximum Likelihood Estimation (MLE) to estimate the mixing coefficients:

  $$P(\theta|y) = \frac{P(y|\theta) \cdot P(\theta)}{P(y)}$$

  where $P(\theta|y)$ is the posterior probability of the sources, θ are the model parameters, and $P(y)$ is the evidence.
- **Empirical Mode Decomposition (EMD)**: EMD decomposes a signal into intrinsic mode functions (IMFs), each representing different frequency components. The decomposition can be formulated as:

  $$x(t) = \sum_{i=1}^{N} IMF_i(t) + \text{Residual}(t)$$

  where $x(t)$ is the original signal, and $IMF_i(t)$ are the intrinsic mode functions. This decomposition allows for separating components associated with different sources by analyzing the IMFs.

E. Challenges of Traditional Techniques

While traditional SCBSS methods have laid the groundwork for separating source signals, they face several key challenges:

- **Limited Scalability**: Many of these techniques, such as ICA and NMF, struggle when the number of sources exceeds the number of available sensors, leading to underdetermined mixtures that are difficult to handle.
- **Linear Mixing Assumption**: Most traditional methods rely on the assumption that the sources are linearly mixed, which may not hold in real-world scenarios where nonlinear mixing occurs.
- **Independence Assumptions**: Techniques like ICA assume that the sources are statistically independent, which may not always be the case in practical situations where sources are correlated.
- **Computational Complexity**: Some traditional methods, such as SCA and Bayesian inference, can be computationally expensive, especially when dealing with large datasets or real-time applications.

Despite these challenges, traditional techniques continue to provide valuable insights into the SCBSS problem and serve as the foundation for more recent advances in the field.

IV. Deep Learning Approaches for SCBSS

In recent years, deep learning has revolutionized the field of signal processing, and its application to Single-Channel Blind Source Separation (SCBSS) has shown remarkable improvements in source separation performance. [1] Traditional methods, such as Independent Component Analysis (ICA) and Non-Negative Matrix Factorization (NMF), rely on statistical assumptions about the sources and their mixing processes. However, deep learning techniques can learn complex features from data, making them highly effective for SCBSS tasks, especially in challenging environments with noise and complex signal mixtures. Deep learning has dramatically advanced the field of SCBSS by offering more flexible, powerful, and accurate models for separating mixed signals. Convolutional networks, recurrent networks, and complex-valued networks have all shown great promise in addressing the challenges of source separation. End-to-end learning and the ability to operate directly on raw signals have further enhanced the performance of deep learning models, making them a compelling choice for SCBSS tasks. However, challenges related to computational complexity, generalization, and source correlation remain, and ongoing research will continue to drive innovation in this area.

A. Convolutional Neural Networks (CNNs) for SCBSS

Convolutional Neural Networks (CNNs) have proven to be particularly effective in SCBSS, as they can capture spatial and temporal features in the time-frequency domain. CNNs have been used in SCBSS to process spectrograms of mixed signals [12] and learn the mapping from mixed signals to their source components. A key advantage of CNNs is their ability to handle high-dimensional data efficiently through the use of convolutional layers that focus on local feature extraction.

In the context of SCBSS, CNNs are typically employed in combination with time-frequency representations of the mixed signals, such as spectrograms, [12] where the network learns to discriminate between the sources by modeling their spectral patterns. These models often include architectures designed for time-series data, making them highly effective in separating sources that overlap in both time and frequency.

B. Recurrent Neural Networks (RNNs) for SCBSS

Recurrent Neural Networks (RNNs) are designed to handle sequential data, making them particularly well-suited for tasks where the order and temporal dynamics of the signals play a critical role. In SCBSS, RNNs are used to model the temporal dependencies of the sources over time. While CNNs excel at capturing spatial patterns, RNNs are effective at learning the sequential nature of time-series data, such as audio signals. [8]

Long Short-Term Memory (LSTM) networks, [7] a type of RNN, have been widely used in SCBSS due to their ability to handle long-term dependencies and avoid issues like vanishing gradients. LSTM-based models can be trained to perform source separation by learning the temporal evolution of each source and how they contribute to the observed mixture. [7] These models have demonstrated significant improvements in the separation of speech signals, especially in scenarios with overlapping speech or background noise.

C. Complex-Valued Neural Networks for SCBSS

Many real-world signal mixtures, particularly in communication systems, involve complex-valued signals, where each signal has both real and imaginary components. Traditional deep learning models typically operate on real-valued data, but complex-valued signals are common in fields such as wireless communications and radar systems. To address this, researchers have developed complex-valued deep learning networks that directly operate on the in-phase and quadrature (I/Q) components of complex signals.

Complex Time-Domain Dilated Convolutional Recurrent Networks (CTDCRN) are a recent development in this area. These networks are designed to process complex signals directly, avoiding the need for transformations to real-valued representations. CTDCRNs combine the strengths of dilated convolutions and recurrent layers to effectively separate sources in the complex domain. They have shown superior performance in scenarios involving complex communication signals, outperforming traditional methods and real-valued networks in terms of separation quality and computational efficiency.

D. End-to-End Learning for SCBSS

One of the most significant advantages of deep learning in SCBSS is the ability to perform end-to-end learning, where the model is trained to directly map the mixed signals to the separated sources without relying on handcrafted features or intermediate steps. This end-to-end approach eliminates the need for pre-processing, feature extraction, and manual tuning, making the model more flexible and adaptable to different types of signal mixtures.

Deep learning models such as Convolutional Time-Domain Audio Separation Networks (Conv-TasNet) have been introduced to provide end-to-end solutions for SCBSS. Conv-TasNet uses a fully convolutional architecture that operates directly on raw audio waveforms, making it possible to separate sources in the time domain without transforming the signals into time-frequency representations. The end-to-end learning process allows Conv-TasNet to learn both the separation and reconstruction of the sources, leading to improved separation quality and reduced computational complexity.

E. Training Deep Learning Models for SCBSS

Training deep learning models for SCBSS requires large and diverse datasets to ensure the model generalizes well to real-world scenarios. In practice, training datasets often consist of mixtures of sources with varying SNRs, overlapping sources, and different types of interference. To effectively train deep learning models, it is crucial to have well-labeled datasets that contain examples of mixed signals and their corresponding source signals.

In addition to supervised learning, unsupervised and semi-supervised learning approaches have also been explored for SCBSS. Unsupervised learning methods aim to learn the separation task without relying on labeled data, which is especially useful in scenarios where obtaining labeled datasets is challenging. These approaches typically involve the use of clustering or generative models to learn the underlying structure of the data and separate the sources.

F. Challenges and Future Directions

Despite the impressive progress made in applying deep learning to SCBSS, several challenges remain. One of the key challenges is the computational complexity of deep learning models, especially when dealing with large datasets and high-dimensional signals. Optimizing the performance of these models while maintaining efficiency is an ongoing area of research.

Another challenge is the generalization of deep learning models to unseen mixtures or environments. Deep learning models are often trained on specific datasets and may not perform well when exposed to new, unseen signal mixtures. Techniques such as domain adaptation, transfer learning, and data augmentation are being explored to address this issue and improve the generalization ability of SCBSS models.

Finally, while deep learning techniques have significantly improved SCBSS performance, they still face limitations in separating highly correlated sources or dealing with very low SNR conditions. Hybrid models that combine the strengths of traditional signal processing techniques with deep learning could provide more robust and reliable solutions in such scenarios.

V. Hybrid Models and Emerging Techniques

While traditional methods and deep learning approaches have individually made significant contributions to the advancement of single-channel blind source separation (SCBSS), recent research has explored the synergy between them to further enhance separation performance. Hybrid models that combine the interpretability and structure of classical techniques with the flexibility and learning capacity of neural networks [10] offer promising solutions to the complex challenges of SCBSS.

A. Traditional-Inspired Deep Models

Hybrid architectures integrate traditional signal processing principles into deep learning frameworks to achieve more robust and generalizable models. For instance, Independent Component Analysis (ICA) has been combined with Convolutional Neural Networks (CNNs) to first decorrelate signal mixtures and then perform feature extraction and separation. Similarly, Non-negative Matrix Factorization (NMF)-guided Deep Neural Networks (DNNs) use learned NMF bases to constrain or initialize the network parameters, thus improving convergence and interpretability.

These hybrid designs often benefit from prior domain knowledge and allow better control over model behavior, especially in underdetermined or noisy scenarios. Moreover, techniques such as Sparse Coding or Empirical Mode Decomposition (EMD) can be used as pre-processing steps to reduce input complexity before feeding the data into a neural model.

B. Transformer-Based Architectures

Transformers have revolutionized natural language processing and are now gaining traction in the signal processing domain, including SCBSS. Unlike RNNs, which struggle with long-term dependencies, transformers utilize self-attention mechanisms to model global temporal relationships in signals.

Separation Transformer (SepFormer), for example, has demonstrated state-of-the-art results in speech separation by leveraging a dual-path transformer network that alternates between intra-chunk and inter-chunk attention. This enables the model to focus on both local and global contextual information, leading to high-quality separation even in overlapping speech or noisy environments.

C. Attention Mechanisms in SCBSS

Attention mechanisms have also been embedded into CNN and RNN architectures to dynamically focus on the most relevant time-frequency regions of the input signal. These attention layers can be learned end-to-end and allow the model to weigh informative segments more heavily during training.

Channel attention modules, for instance, have been used in convolutional time-domain networks to emphasize low-level features critical for waveform reconstruction. Similarly, time-domain audio separation models [9] such as Conv-TasNet have been extended with temporal attention to improve separation under variable SNR conditions.

D. Diffusion Models for Source Separation

Diffusion probabilistic models, originally developed for generative image modeling, have recently been applied to audio source separation. These models learn the data distribution through a gradual denoising process and can generate clean source signals from noisy or mixed inputs.

In SCBSS, diffusion models are explored to iteratively estimate the posterior distribution of clean sources, offering a promising alternative to traditional deterministic mappings. Early studies show that these models can outperform conventional neural networks in terms of perceptual quality and robustness, especially in scenarios with severe noise or interference.

E. Future Outlook

The integration of traditional signal processing with emerging AI techniques is expected to continue evolving. As transformer-based and diffusion models mature, they may redefine the performance ceiling of SCBSS, particularly when combined with domain knowledge and hybrid learning paradigms. The challenge lies in balancing model complexity, interpretability, and real-time applicability, paving the way for robust and scalable separation systems across various domains.

VI. Evaluation Metrics and Benchmarks

To quantitatively assess the performance of Single-Channel Blind Source Separation (SCBSS) systems [6] , a variety of metrics have been developed. These metrics evaluate the quality of separation from multiple perspectives such as distortion, interference suppression, artifacts, and perceptual quality. This section outlines the most commonly used metrics and benchmark datasets in the field.

A. Evaluation Metrics

1) Signal-to-Distortion Ratio (SDR): The Signal-to-Distortion Ratio (SDR) is a widely used metric that evaluates the overall quality of the separated signal compared to the ground truth. It is defined as:

$$\mathrm{SDR} = 10 \log_{10} \left(\frac{\|s_{\text{target}}\|^2}{\|e_{\text{total}}\|^2} \right) \tag{2}$$

where s_{target} is the projection of the estimated signal onto the reference signal, and e_{total} is the total error including noise, interference, and artifacts.

2) Signal-to-Interference Ratio (SIR): SIR measures how well the algorithm suppresses interference from other sources. It is calculated as:

$$\mathrm{SIR} = 10 \log_{10} \left(\frac{\|s_{\text{target}}\|^2}{\|e_{\text{interf}}\|^2} \right) \tag{3}$$

where e_{interf} is the part of the error due to interference from other sources.

3) Signal-to-Artifact Ratio (SAR): SAR assesses the amount of artifacts introduced during the separation process:

$$\mathrm{SAR} = 10 \log_{10} \left(\frac{\|s_{\text{target}} + e_{\text{interf}}\|^2}{\|e_{\text{artif}}\|^2} \right) \tag{4}$$

where e_{artif} is the component of the error attributed to algorithmic artifacts.

4) Bit Error Rate (BER): In wireless communication applications, the Bit Error Rate (BER) is a key metric. It is defined as the ratio of the number of incorrectly received bits to the total number of transmitted bits:

$$\mathrm{BER} = \frac{N_{\text{error}}}{N_{\text{total}}} \tag{5}$$

where N_{error} is the number of erroneous bits, and N_{total} is the total number of transmitted bits.

5) Perceptual Evaluation of Speech Quality (PESQ): PESQ is an objective metric used to estimate the perceived quality of speech. It compares the original and degraded signals using a perceptual model and outputs a score between 1 (bad) and 5 (excellent). Although the full PESQ model is proprietary, it is widely used in voice quality evaluation for telecommunication systems.

B. Benchmark Datasets

The following benchmark datasets are commonly used for training and evaluating SCBSS models:

- **WSJ0-2mix:** A dataset based on the Wall Street Journal corpus, widely used for speech separation. [5] It contains mixtures of two speakers and their corresponding clean sources.
- **MIR-1K:** Contains 1,000 song clips sampled at 16 kHz, mainly used for singing voice separation. Each clip includes a clean vocal and accompanying music.
- **CHiME:** Designed for robust ASR, CHiME datasets contain real and simulated noisy speech recorded in everyday environments.
- **LibriMix:** Derived from the LibriSpeech corpus, LibriMix provides mixtures of clean speech for source separation tasks, with adjustable SNR levels and speaker count.

These datasets help standardize evaluations and enable fair comparisons among various SCBSS algorithms.

VII. Single-Channel Blind Source Separation in Image Signals

Single-Channel Blind Source Separation (SCBSS) techniques, traditionally applied to audio and communication signals, have found applications in the realm of image processing. The objective in this context is to decompose a single observed image into its constituent source images without prior information about the mixing process or the sources themselves. This problem is particularly pertinent in scenarios where multiple images overlap or are superimposed, and the goal is to recover the original images from the composite observation.

A. Applications in Image Processing

In image processing, SCBSS is utilized in various applications, including:

- **Document Analysis:** Separating overlapping text and background images to enhance readability and facilitate optical character recognition.
- **Medical Imaging:** Isolating different tissue types or highlighting specific features from composite medical images.
- **Astronomical Imaging:** Extracting distinct celestial objects from composite space observations where multiple sources overlap.

B. Methodologies

The methodologies for SCBSS in image signals often draw parallels with those in audio signal processing but are adapted to the two-dimensional nature of images. Traditional approaches have been complemented in recent years by powerful AI-based models capable of learning complex, non-linear transformations.

- **Independent Component Analysis (ICA):** Assumes that the source images are statistically independent and aims to find a linear transformation that minimizes the statistical dependence between the separated components.
- **Non-negative Matrix Factorization (NMF):** Decomposes the observed image into non-negative components, suitable for images where pixel values represent quantities that cannot be negative.
- **Sparse Component Analysis (SCA):** Leverages the sparsity of source images in a suitable transform domain to achieve separation.
- **Convolutional Neural Networks (CNNs):** CNNs are used for image-based SCBSS by learning spatial hierarchies in image data, making them suitable for local pattern extraction and image denoising tasks.
- **Autoencoders:** Typically used in unsupervised learning, autoencoders learn compressed latent representations of images. In the context of SCBSS, they can be trained to reconstruct individual source images from mixed inputs, either in an unsupervised manner or with supervision when clean targets are available.
- **Generative Adversarial Networks (GANs):** GANs can be used for image separation by learning to generate individual source images that are statistically similar to

true source distributions. Conditional GANs (cGANs) are particularly useful when auxiliary information is available.

- **Transformer-based Vision Models:** Recent advances also include the application of Vision Transformers (ViTs) that process image patches and model long-range dependencies, offering promising results in separating spatially entangled features.

These AI-based methods excel at learning non-linear and hierarchical representations, making them highly effective for complex image separation tasks where traditional linear methods often fall short. Moreover, they benefit significantly from large-scale labeled datasets such as MNIST, EMNIST, and Fashion-MNIST for supervised learning, or use synthetic mixtures for unsupervised and self-supervised training.

C. Benchmark Datasets

To evaluate and benchmark SCBSS algorithms in image processing, several datasets are commonly employed:

- **MNIST:** The Modified National Institute of Standards and Technology database is a large collection of 28x28 grayscale images of handwritten digits ranging from 0 to 9. It comprises 60,000 training images and 10,000 test images. MNIST serves as a standard benchmark for image processing algorithms.
- **EMNIST:** The Extended MNIST dataset is an expansion of MNIST, including both handwritten digits and letters. It offers a more comprehensive set of 28x28 grayscale images, facilitating research in character recognition across a broader spectrum.
- **Fashion-MNIST:** This dataset consists of 28x28 grayscale images of fashion items across 10 categories, such as shirts, trousers, and shoes. Fashion-MNIST was developed as a more challenging drop-in replacement for the original MNIST dataset, providing a benchmark for machine learning algorithms in the context of fashion product recognition .
- **CIFAR-10:** Comprising 60,000 32x32 color images across 10 classes, including animals and vehicles, CIFAR-10 is widely used for evaluating image recognition and classification algorithms .
- **ImageNet:** A large-scale dataset containing millions of annotated images across thousands of categories, ImageNet serves as a comprehensive benchmark for image classification and object detection tasks .

These datasets provide standardized benchmarks that facilitate the evaluation and comparison of SCBSS algorithms in image processing applications.

VIII. Comparison of SCBSS Solutions

A comparative evaluation of Single-Channel Blind Source Separation (SCBSS) methods reveals distinct trade-offs between traditional statistical approaches, modern deep learning techniques, and emerging hybrid models. Table 1 summarizes the key characteristics across these paradigms. The insights are discussed below based on recent findings in the literature.

TABLE I
Comparison of SCBSS Methods

Criteria	Traditional	Deep Learning	Hybrid
Assumption	Statistical independence, linear mixing	Learns mappings from data	Combines priors with learning
Model Complexity	Low to moderate	High (GPU needed)	Moderate to high
Accuracy (SDR/SIR)	Limited in noise and overlap	High with large data	Competitive, more robust in underdetermined settings
Generalization	Poor to moderate	May overfit; domain sensitive	Better via priors and structure
Training Data	Minimal or none	Requires large labeled sets	May reduce data needs
Real-Time Use	High efficiency	Often too heavy	Improving with light models
Application	Audio, speech	Speech, bio-signals, communication	Low-SNR, wireless, complex mixes

A. Assumptions and Model Structure

Traditional techniques such as Independent Component Analysis (ICA) and Non-Negative Matrix Factorization (NMF) operate under strong assumptions like statistical independence and linear mixing . While this offers mathematical elegance, it limits their applicability in real-world, non-linear environments.

In contrast, deep learning methods including CNNs , RNNs, and CTDCRNs learn feature representations directly from data without requiring prior assumptions. Hybrid models strike a balance by incorporating domain-specific priors into neural frameworks, enhancing interpretability and stability .

B. Complexity and Real-Time Use

Traditional models remain computationally efficient and suitable for real-time use, especially when sensor or compute constraints are present. Deep learning models, while accurate, often require high-performance GPUs due to their large parameter sets. However, recent architectures like Conv-TasNet and CNSE networks have proposed optimizations that reduce latency and memory usage .

C. Accuracy and Robustness

Deep neural networks have shown high performance in terms of Signal-to-Distortion Ratio (SDR) and Signal-to-Interference Ratio (SIR), particularly under noisy and overlapping conditions . Hybrid models also demonstrate competitive accuracy by leveraging structure-aware learning strategies .

D. Generalization and Data Requirements

A major strength of traditional methods is their ability to work with minimal or no training data. Deep models, while powerful, tend to be data-hungry and can overfit to domain-specific datasets. Hybrid models aim to reduce data dependency by using handcrafted priors and transfer learning techniques .

E. Application and Scalability

SCBSS techniques find application across various domains from audio and speech separation to wireless communications. [3] Complex-valued networks like CTDCRN have shown notable improvements in communication signal recovery . Meanwhile, hybrid and transformer-based methods are extending applicability to underdetermined, low-SNR, and dynamic environments

In summary, while no single SCBSS solution is universally optimal, deep learning models currently offer the best separation quality. Hybrid methods present a promising direction by balancing accuracy, interpretability, and computational feasibility.

IX. Application Domains

Single-Channel Blind Source Separation (SCBSS) is a cornerstone technique in signal processing with a wide spectrum of real-world applications. The ability to separate multiple source signals from a single observed mixture without prior knowledge of the sources or mixing process has made SCBSS invaluable across various domains.

A. Speech and Audio Processing

Perhaps the most iconic illustration of SCBSS is the *cocktail party problem*, where a listener focuses on one speaker amidst a crowd. SCBSS enables automated systems to mimic this ability, facilitating robust speech recognition, transcription, and enhancement. Applications include virtual assistants, hearing aids, and mobile telephony. Time-frequency masking methods [4] and deep learning-based architectures such as CNNs and RNNs have significantly improved separation quality in these settings. [8]

B. Wireless Communications

In wireless systems, [3] co-frequency signal separation is crucial for improving spectrum efficiency. [2] SCBSS plays a vital role in separating modulated communication signals that share the same channel. Recent works has introduced complex-valued neural networks, such as the Complex Time-Domain Dilated Convolutional Recurrent Network (CTDCRN), which directly processes in-phase and quadrature (I/Q) components for signal recovery. These approaches have proven effective in minimizing bit error rates (BER) and optimizing performance in low signal-to-noise ratio (SNR) environments.

C. Biomedical Signal Processing

Biomedical applications leverage SCBSS to extract meaningful signals from noisy recordings. For instance, fetal ECG or EEG signals are often contaminated with maternal and environmental noise. SCBSS methods can isolate the weak, target signals from these mixtures, enhancing diagnostic accuracy in non-invasive procedures.

D. Acoustic Scene Analysis and Surveillance

SCBSS is employed in environmental monitoring systems to separate sound events from urban noise, enabling automated event detection. For example, isolating gunshots, alarms, or human voices in surveillance footage benefits from advanced SCBSS models trained on diverse acoustic scenes.

E. Image and Vision Systems

Though traditionally focused on audio, SCBSS techniques have been extended to single-channel image decomposition. Applications include separating overlapping text in document analysis, isolating structures in medical imaging, and extracting celestial bodies in astronomical images. Models such as CNNs, autoencoders, [11] and even transformer-based vision architectures have been adapted for spatial domain signal separation.

F. Multimedia and Entertainment

In multimedia applications, SCBSS helps separate music sources, e.g., extracting vocals, drums, or instruments from a mixed track. This has applications in karaoke systems, music remixing, and restoration of archival recordings. Deep learning models trained on large audio datasets are particularly effective here.

G. Emerging Domains

Emerging areas such as Internet-of-Things (IoT), robotics, and brain-computer interfaces (BCIs) increasingly depend on robust source separation in dynamic, single-channel contexts. Low-latency and real-time capabilities of SCBSS models are being explored for embedded systems, autonomous vehicles, and wearable health devices.

In general, the flexibility and adaptability of SCBSS methods, especially those augmented by artificial intelligence, are opening doors to new applications, highlighting the growing relevance of the technique across disciplines.

X. Research Gaps and Future Directions

Despite significant advancements in Single-Channel Blind Source Separation (SCBSS), several challenges persist, paving the way for future research avenues.

A. Handling Nonlinear and Nonstationary Mixtures

Most existing SCBSS methods assume linear and stationary mixing processes. However, real-world scenarios often involve nonlinear and nonstationary mixtures. Developing models that can effectively address these complexities remains an open challenge. Recent approaches, such as multi-encoder autoencoders, [11] have shown promise in tackling nonlinear mixtures, but further exploration is needed to enhance their applicability and performance.

B. Generalization Across Diverse Signal Types

Current SCBSS techniques often specialize in specific domains, such as audio or biomedical signals. Achieving models that generalize well across diverse signal types without extensive retraining is a significant research gap. Investigating domain-agnostic architectures and training paradigms could lead to more versatile SCBSS solutions.

C. Integration of Advanced Deep Learning Architectures

The incorporation of advanced deep learning architectures, like transformers, has shown potential in improving separation performance. For instance, transformer-guided GANs have been applied to image separation tasks. Extending such architectures to other SCBSS applications and understanding their benefits and limitations is a promising direction for future research.

D. Robustness to Noise and Interference

Ensuring that SCBSS methods are robust to various noise levels and types of interference is crucial for practical applications. Developing algorithms that maintain high separation quality in adverse conditions remains an ongoing challenge.

E. Real-Time Processing Capabilities

Many SCBSS applications, such as speech enhancement in communication systems, require real-time processing. Designing computationally efficient algorithms that can operate with low latency is essential for these time-sensitive applications.

F. Evaluation Metrics and Benchmarking

There is a need for standardized evaluation metrics and comprehensive benchmarking datasets to facilitate the objective comparison of SCBSS methods. Establishing such standards would aid in assessing progress and identifying the most promising approaches.

Addressing these research gaps will contribute to the development of more robust, efficient, and versatile SCBSS methods, expanding their applicability across various domains and real-world scenarios.

XI. Conclusion

Single-Channel Blind Source Separation (SCBSS) remains a critical and evolving area within the field of signal processing. Over the years, it has transformed from relying solely on traditional statistical models, such as Independent Component Analysis (ICA) and Non-negative Matrix Factorization (NMF), to integrating advanced machine learning techniques, especially deep learning architectures like CNNs, RNNs, and more recently, transformers and complex-valued neural networks.

This survey has reviewed the fundamental principles, classical approaches, and the significant leap made possible through AI-driven models in SCBSS. Deep learning has introduced powerful tools for capturing non-linear dependencies and performing end-to-end source separation from raw inputs. Furthermore, hybrid methods combining conventional signal processing with modern AI paradigms offer promising improvements in robustness, accuracy, and interpretability.

Despite these advancements, SCBSS still faces several open challenges. Generalization across diverse signal domains, robustness under low SNR and non-stationary conditions, and real-time deployment in constrained environments remain active areas of research. Moreover, the lack of standardized evaluation protocols and domain-agnostic architectures continues to hinder direct comparison and practical adoption of SCBSS models.

Future research should focus on expanding model adaptability through semi-supervised learning, domain transfer techniques, and physics-informed networks. The exploration of transformer-based and diffusion-inspired architectures also holds potential for reshaping how separation is approached, especially in highly complex mixtures.

In summary, SCBSS is no longer a purely academic pursuit: it is a foundational enabler across audio processing, wireless communications [3] biomedical diagnostics, surveillance, and emerging IoT ecosystems. With continued interdisciplinary collaboration and innovation in AI, the field is poised for breakthroughs that will drive intelligent signal processing in the next generation of real-world systems.

References

[1] S. Ansari, A. S. Alatrany, K. A. Alnajjar, T. Khater, S. Mahmoud, D. Al-Jumeily, and A. J. Hussain, "A survey of artificial intelligence approaches in blind source separation," *Neurocomputing*, vol. 561, p. 126895, 2023.

[2] X. Hou and Y. Gao, "Single-channel blind separation of co-frequency signals based on convolutional network," *Digital Signal Processing*, vol. 129, p. 103654, 2022.

[3] P. Guo, M. Yu, L. Shen, Z. Lin, K. An, and J. Wang, "Single-channel blind source separation in wireless communications: A complex-domain deep learning approach," *IEEE Wireless Communications Letters*, vol. 13, no. 6, pp. 1645–1648, Jun. 2024.

[4] A. J. R. Simpson, "Probabilistic binary-mask cocktail-party source separation in a convolutional deep neural network," *arXiv preprint arXiv:1502.01272*, 2015.

[5] D. L. Wang and J. Chen, "Supervised speech separation based on deep learning: An overview," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, doi:10.1109/TASLP.2018.2842159, 2018.

[6] K. Patki, "Review of single channel source separation techniques," in *Proc. ISMIR*, 2013.

[7] M. Zhao, X. Yao, J. Wang, Y. Yan, X. Gao, and Y. Fan, "Single-channel blind source separation of spatial aliasing signal based on stacked-LSTM," *Sensors*, vol. 21, no. 14, p. 4844, 2021, doi:10.3390/s21144844.

[8] J. He, W. Chen, and Y. Song, "Single Channel Blind Source Separation Under Deep Recurrent Neural Network," *Wireless Personal Communications*, vol. 114, pp. 2245–2263, 2020, doi:10.1007/s11277-020-07624-4.

[9] M. T. S. Al-Kaltakchi, A. S. Mohammad, and W. L. Woo, "Ensemble System of Deep Neural Networks for Single-Channel Audio Separation," *Information*, vol. 14, no. 7, p. 352, 2023, doi:10.3390/info14070352.

[10] S. Liu, B. Wang, and L. Zhang, "Blind Source Separation Method Based on Neural Network with Bias Term and Maximum Likelihood Estimation Criterion," *Sensors*, vol. 21, no. 3, p. 973, 2021, doi:10.3390/s21030973.

[11] M. B. Webster and J. Lee, "Blind Source Separation of Single-Channel Mixtures via Multi-Encoder Autoencoders," *arXiv preprint arXiv:2309.07138*, 2023.

[12] D. Michelsanti, Z.-H. Tan, S.-X. Zhang, Y. Xu, M. Yu, D. Yu, and J. Jensen, "An Overview of Deep-Learning-Based Audio-Visual Speech Enhancement and Separation," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 29, pp. 102–122, 2021.