EasyChair Preprint

№ 9915

Unsupervised Hidden State Estimation and Blind Source Separation Using Auto-Encoder RNN Filter

Clint Steed and Kim Namhun

March 31, 2023

Unsupervised hidden state estimation and blind source separation using Auto-encoder RNN

1st Clint Alex Steed
Department of Industrial Engineering
Stellenbosch University
Stellenbosch, South Africa
casteed@ieee.org

2nd Namhun Kim
Department of Mechanical Engineering
Ulsan National Institute of Science and Technology
Ulsan, South Korea
nhkim@unist.ac.kr

***Abstract*— Human internal state affects operator well-being and production outputs, but it cannot be directly measured and must be estimated. This paper proposes a deep learning approach to unsupervised nonlinear hidden state estimation using an auto-encoder, by framing it as a blind source separation (BSS) problem. The model is composed of an auto-encoder-based recurrent neural network (RNN) and extended to blind source separation through the use of local losses to decorrelate hidden signals. The number of sources can be determined by adjusting the dimension of the hidden state signal. Simulations demonstrate hidden state extraction when the correct dimensionality is selected and separation of multiple sources. Using an auto-encoder in the model restricts it to cases where there are more sensors than hidden states. This makes it well-suited for domains with redundant sensors, such as drones and self-driving cars.**

Keywords—Unsupervised learning, Hidden State Estimation, Recurrent Neural Networks, Auto-Encoders, Blind source separation.

I. Introduction

Human-centric systems are actively being researched, as evidenced by several recent special issues [1]–[3], an EU report [4], and the rising cost of labor in manufacturing. This requirement, along with rapid change in systems encourages the use of automated/unsupervised approaches. However, these approaches are limited in their application of human control due to ethical issues.

When optimizing human operation systems we are presented with the challenge that “Systems Serve Humans”. Therefore, these systems must balance production requirements with the well being of operators. The authors believe this is possible by controlling systems based on human internal state variables like fatigue.

This work is the starting point of investigating automatically extracting the human internal state. The problem is formulated as a blind source separation problem. A nonlinear deep estimator is developed to extract the unmeasurable states. The main objective of the current investigation is to develop a model that meets the requirement to extract the human hidden states. Capable of time-varying and state interaction, or simply nonlinear dynamic.

The contributions are:

1. Development of a dynamic nonlinear model with time-varying and source interaction capabilities.
2. By formulating image generation as a BSS problem, we gain cross pollination of methods, leading to insights that motivate the models development.

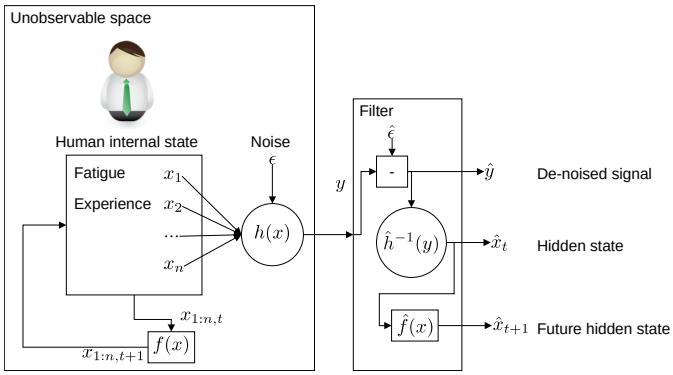

Figure 1: Human state estimation as a dynamic problem showing estimator requirements

II. Literature

A. The importance of human unmeasurable states

States like fatigue cannot be measured directly, instead they are estimated by measuring their effects. Sensing can be accomplished either by gathering data from the operator or by measuring the impact on production signals.

1) Operator sensing

Industries with static operators (pilots, long distance drivers) have seen successful commercial products using operator facing cameras and integrating sensors into instrumentation.

However, industries with dynamic operator tasks (manufacturing and seafarers) have not enjoyed the same success. Wearable sensors like accelerometers, EMGs, and temperature sensors [10] have been used in lab experiments, but hinder operator comfort. Biological samples like oral swabs [11] are accurate, but not suitable for in-situ sensing.

The issue is that many of these data acquisition methods are not feasible for in-situ sensing and it is not clear whether the information provided overlaps.

2) Production data signals

Fatigue negatively impacts production outcomes but does not provide information about the underlying causes. However, it can be used as an indicator to estimate the human state.

For instance, factors such as time of day and consecutive work days are strong indicators of risk of injury [5], [6]. This hints that there are multiple modalities to fatigue. We expect one source for daily fatigue and another for weekly fatigue. Another example is that learning increases operator’s throughput rate, while circadian rhythms [7], forgetting [8], work-rest ratios [9] decrease it.

Human state estimation is important for human well being as it can reduce risk of injury and production because it can affect production outputs. The examples illustrate the model should be (1) dynamic, allowing time varying signals and (2) nonlinear, allowing hidden state/source interaction.

B. Blind source separation

The blind source separation problem is sometimes better described by the cocktail party problem. Imagine numerous people talking, resulting in the recipient receiving a mixed sound signal and having to discern between different conversations. The authors use this term as a problem formulation rather than a collection of methods. BSS methods have been used for audio source separation [12] and signal processing [13].

BSS is often an ill conditioned problem, resulting in numerous solutions. Specifying further constraints has the potential to reduce this. In some cases, the signal can be recovered but not the amplitude. However, this limitation can be overcome by using a number of local losses, which will be discussed shortly.

One technique for achieving BSS is Independent Component Analysis (ICA), an extension to the well known Principal Component Analysis (PCA).

C. Deep blind source separation as high level feature separation

The authors argue that deep learning makes it possible to represent several high-level tasks as a blind source separation (BSS) problem. Among these, image processing and generation techniques are the easiest to visualize. This insight is valuable for conditioning the latent space, as high-level feature modification is often necessary to generate new images.

For instance, [14] separates facial identities from emotions to reconstruct faces with different emotions, [15] separates blur, noise, and compression image distortions, and Fader networks [16] allow sliding attributes to adjust the feature intensity, such as transforming from young to old.

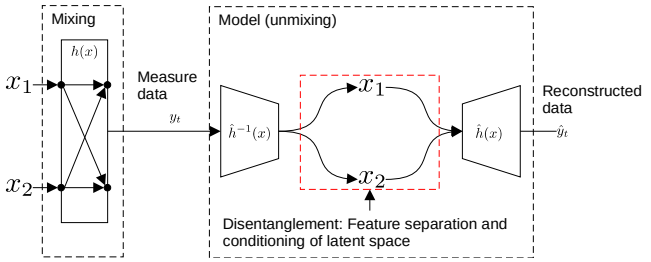

Figure 2: Deep blind source separation using an auto-encoder showing disentanglement of the latent space

Disentanglement refers to the linear separation of features, styles, and other information in the latent space [17], [18], which is similar to source separation. Several methods have been used for estimating disentanglement such as, developing a linear classifier [19], using cluster separability [20], applying a probabilistic total correlation penalty that requires sampling [21], or using a discriminator [16]. While auto-encoders are commonly used for this problem [17], [22], [23], some models do not employ them [18], making disentanglement comparison difficult.

D. Hebbian learning inspired local losses

One neural learning algorithm which has shown lots of promise in this area is Hebb learning. Although it is not used in this work due to back-propagation tools being more mature. The insights found in Hebb learning motivate the choices for local losses here.

Hebbian learning is best described by the adage “Neurons that fire together, wire together” [24] . The Hebbian learning interpretation of this strengthens of pre-synaptic and post-synaptic pairs that fire together. This results in learning the principal components [25]. On the other hand, Anti-Hebbian learning weakens pre-synaptic and post-synaptic pairs that don’t fire together, resulting in decorrelation which can be used for BSS [26], [27]. This Anti-Hebbian learning can be imitated using an auto-encoder with the inappropriately named Decov loss [14], [28].

Hebbian learning has also addressed some of the other limitations in BSS, by conditioning the source signals. Unscaled source amplitude is addressed by enforcing unit variance [27], this in turn inspires the use of unit variance local loss use here. Similarly, zero mean source signal is typically achieved by whitening the data, instead we use a small zero-mean loss.

E. Auto-encoder implications on sensor design

The auto-encoder is selected as the starting point for the model because there is strong evidence that it performs nonlinear Principal Component Analysis (PCA) [29]. The intuition here is to use decorrelation to move toward nonlinear Independent component analysis (ICA), one of the better known methods for BSS. However, the auto-encoder does place some restrictions on our sensor selection. It assumes that the number of sensor signals is greater than the number of source signals, $m > n$ where $x \in R^n$ $y \in R^m$. This is not unreasonable since redundant low-cost sensors are often preferred over fewer high-cost sensors and provides denoising benefits.

F. Deep temporal estimators

Most BSS work considers static solutions. For example Fourier transform and have the limitation on time varying signals and latent state interaction. Formulating this problem as a dynamical system has the potential to relax these two limitation.

Well known estimators like the Kalman filter and extended Kalman filter have been widely applied. However, their linear limitations are known [30], [31]. Another generation of filters use computationally intensive monte carlo simulations to estimate nonlinear behavior [32]. Deep estimation techniques tend to incur this computational cost upfront by learning filtering parameters and estimating functions, resulting in cost effective inference. Here, the functions or parameters are learned. A desirable trait with deep filters is the ability to include prior known information , usually in the form of partially known dynamics [33].

G. Summary

To summarize, human state estimation can benefit human well being and production output. These models require time varying and interacting source capabilities. Due to the variety of sensing means, humans automated/unsupervised estimation will be beneficial. On the other hand, the ability to incorporate known dynamics is desirable.

Since the human state cannot be measured directly, we suggest modeling it as a BSS problem. By formulating deep image-processing tasks also as a BSS problem, we gain the insight that a decorrelated AE perform nonlinear ICA. Hebbian inspired local losses can address the limitations of ill-conditioned models.

III. Theory

The figure that follows depicts the decisions made when developing the model. Starting from a standard auto-encoder, moving towards a supervised temporal estimator, and then an unsupervised estimator.

We begin by developing the model, then describing the local losses required to shape the latent state.

A. Neural architecture and losses

The model is developed with three losses, starting from a standard auto-encoder with the reconstruction loss $L_1 = ||y - \hat{y}||$ included.

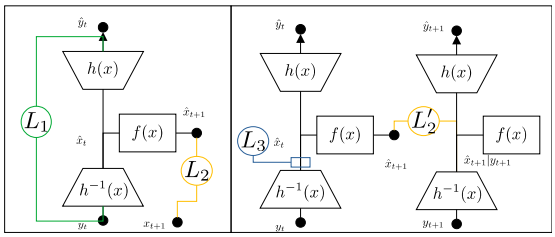

Figure 3: From left (1) shows the supervised estimator and (2) the unsupervised estimator.

Next we create a supervised estimator by adding an evolution/transition function $x_{t+1} = f(x_t)$ and loss $L_2 = ||\hat{h}^{-1}(y_t) - x_{t+1}||$. Figure 2 illustrates this. Some important notes here are that (1) we assume our sensor dimension to be higher than our latent dimension $y \in R^m, x \in R^n, m > n$. This is advantageous since it is common to have numerous redundant sensors for noise reduction, cost reduction, and reliability. At this stage we are presented with a choice. Ideally, the transition function is known and used. This work instead assumes an ANN is used. The hidden state data X is used for training an ANN but in the next step, we lift this condition.

The final step is unrolling in a similar way to other recurrent neural networks. Here a loss penalizes the error between the sequential predictions of the model $L_2' = ||\hat{f}(\hat{h}^{-1}(y_t)) - \hat{h}^{-1}(y_{t+1})||$, specifically the encoded temporal-prediction from the current time $\hat{x}_{t+1}|\hat{x}_t$ and next encoded prediction $\hat{x}_{t+1}|y_{t+1}$. Figure 3 illustrates this. This change removes the requirement for the hidden state data X, relying only on Y returning to an unsupervised learning problem. The cost of this is that batches of at least 2 sequential data-points be used. If the equation $f(x)$ is known, it can be substituted for the neural network.

Since we do not supply the function, the model must infer it, which can result in several effects. Firstly, the dimensionality of x now becomes a design choice, meaning that we can decide how many variables to include in our input. Secondly, this is a poorly conditioned problem, which means that numerous solutions exist and we may not receive the same solution between multiple training sessions. In other words, the model may converge to different solutions each time it is trained.

B. Local losses

A widely accepted strategy to address the issue of numerous solutions is to calibrate the source signals to some domain, for example $x \in [0, 1]$. In this work, we use a number of losses that are local to the mini-batch used in training. First, a mean loss encourages zero mean $L_\mu(x) = |x|$. The second loss ensures unit variance $L_\sigma(x) = |1 - \sigma(x)|$.

A decorrelation loss disentagles sources $L_{DeCov}(x) = ||C|| - ||diag(C)||$, where $C_{i,j} = \frac{1}{N}(x_i, \mu i)(x_j, \mu_j)$. The intuition for this choice is moving from PCA to ICA. The resulting local losses can then be weighted and summed, $L_3(x) = L_\mu + L_\sigma + L_{Decov}$.

C. The dimensionality of the latent space

Given this model, one design choices is to choose the dimensionality of x, n where $x \in R^n$. The dimensionality of y, m where $y \in R^m$, is dictated by the sensors. We will select n such that we can learn more about the system.

In summary, we now have a model that is capable of dynamic estimation of sources with interaction. Although the model can incorporate known dynamics in the form of the transition function, this work is interested in inferring the transition. This introduces the dimension of the latent space as a design parameter.

IV. Methodology

Two simulations are conducted. The first investigates the effect of selecting the dimensionality of the latent space. The second simulation investigates extracting multiple nonlinear sources.

A. Model

In order to evaluate the filters behavior they are tested on a toy problem of one pendulum acting as a single source. The state is generated by the system transition $x_{t+1} = f(x_t)$. The model receives the sensor signal y which is mixed and noise is added according to $y = h(x)$. The goal of the model is to estimate the transition function $\hat{f}(x)$ and the state estimation function $\hat{x} = \hat{h^{-1}}(y)$.

Two mixing strategies are considered. Firstly, independent nonlinear mixing, via $h(x) = \frac{1}{1+|x_i|}$, which tests the models ability to perform nonlinear estimation. Next, a nonlinear combination mixing, $h(x) = \frac{1}{1+|x_i x_j|}$, testing source separation. These sensor models significantly change the signal and do not allow negative values in these simulations. This has an impact on the resulting transition function.

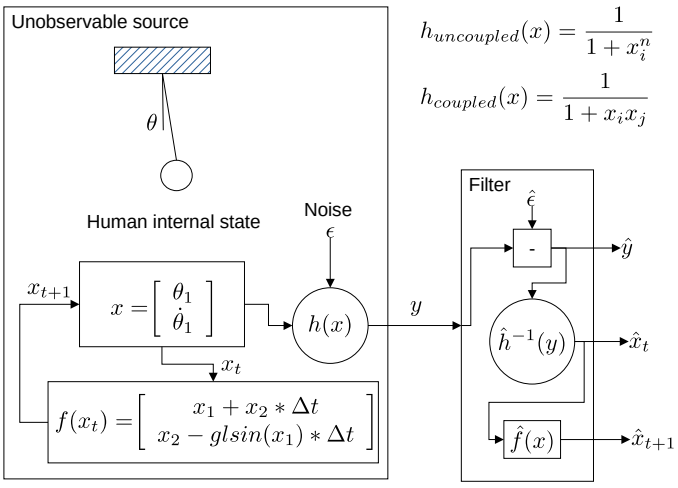

Figure 4: Simulation used for testing model

We explore the selection of the latent space and encounters the repeated signals issue.

B. Multiple sources

Next a system consisting of two sources at different frequencies are used. This section tests source separation.

A number of systems are used. Firstly, the pendulum is selected for its familiarity. Also the Van der pol attractor is selected as it can be tuned to represent nonsymmetric waves [34]. Finally, the triangular wave it used due to its discontinuous nature.

V. Results

A. Single source pendulum state estimation

As expected the model infers principal signals. The leading and lagging relationship between the position and velocity was learned. We also see that noise is present in the result, it is unclear if regularization can improve the results. The relative increasing magnitude is also captured, showing time varying signals are captured.

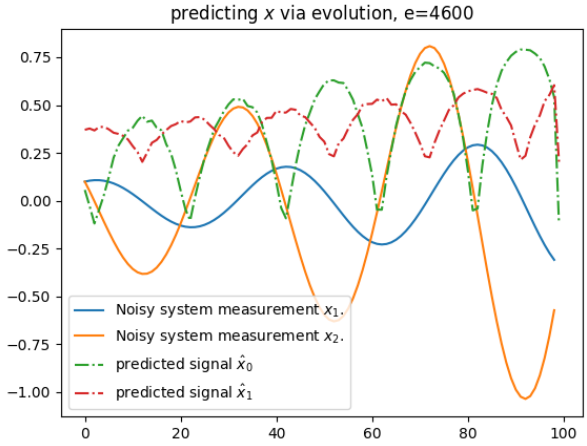

Figure 5: The model successfully estimates the two hidden states (position and velocity) of the pendulum.

This result would indicate that the model is sufficient for decoding and predicting some indicators of the hidden state.

B. Varying the latent space dimensionality

In our experiments, we observed that when the number of source signals (n) is equal to 1, the resulting signal was unique, meaning that different runs with random initialization produced the same output signal. However, when n>1, the results were not unique, and tthe signals' mean and sign would vary. Furthermore, when n was increased beyond 2, repeated signals occurred, which is likely due to the presence of repeated principal components. Therefore, identifying unique signals can be used to select the appropriate principal dimension size for a given dataset. Currently, this process is often done through visual inspection, which can be time-consuming and subjective. Therefore, automated approaches should be investigated to improve the efficiency and reliability of this process.

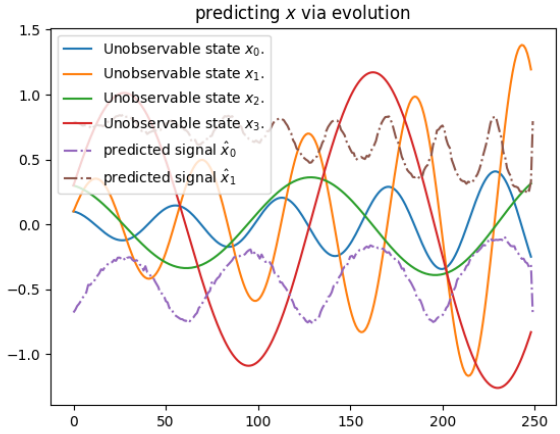

Figure 6: Repeated principal signals are estimated. The authors notice this happens when the dimension of the latent space is too high.

C. Source separation

The process is repeated with multiple sources and decorrelation added to the model to determine whether the model can perform blind source separation.

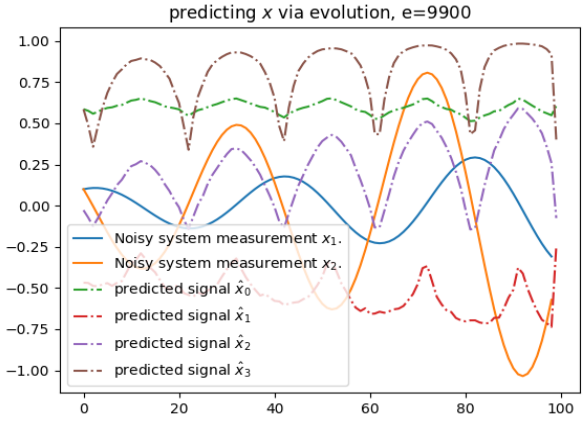

Figure 7: The model performing source separation.

The figure above clearly shows that the separated signals are observed. It is also clear that signals are affected by noise. Again the amplitudes are not repeatable between runs.

D. Common systems

The triangular wave was also reproduced, showing the model can learn signals that are not smooth.

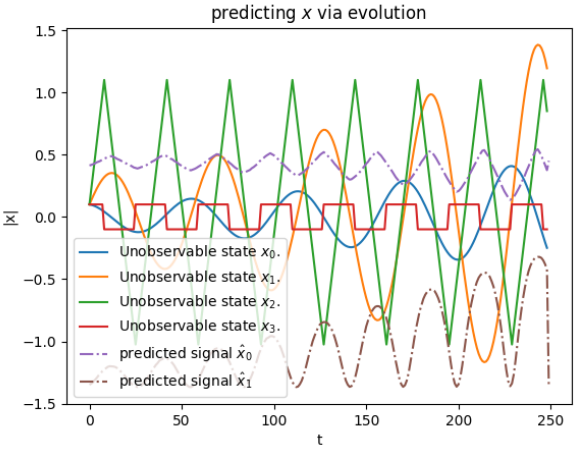

Figure 8: The model separating a triangular wave which has sharp discontinuous peaks.

A Van der Pol attractor was used and the model was able to reconstruct these signals. Showing it can model nonsymmetric waves and limit cycles.

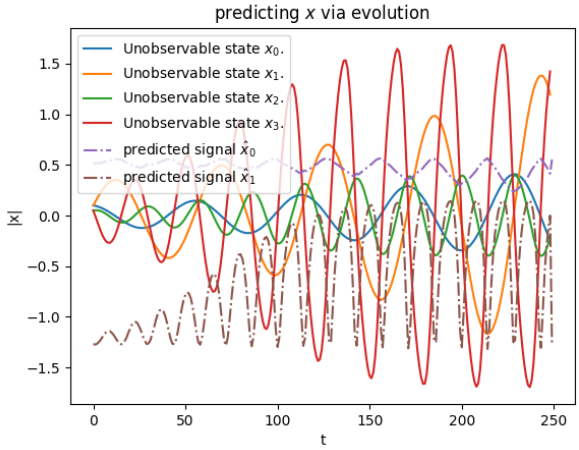

Figure 9: The mode separating a time varying, nonsymmetric Van Der pol attractor.

E. Conclusion and further work

In conclusion, estimating the human state can improve operator well-being and increase production throughput. Using examples from literature, we have determined that the model for human state estimation should be capable of handling nonlinear dynamics. Formulating the task as a blind source separation problem resulted in a deep AE-RNN model that meets these requirements and can infer hidden states in an unsupervised manner, with the option of incorporating prior information in the form of a transition function. The dimensionality of the hidden state was selected based on visual inspection, which is suboptimal and requires further research to develop a more effective selection method.

REFERENCES

[1] N. A. Stanton, "Special issue on human factors and ergonomics methods," *Hum. Factors Ergon. Manuf.*, vol. 32, no. 1, pp. 3–5, 2022, doi: 10.1002/hfm.20943.

[2] F. Sgarbossa, E. Grosse, W. P. Neumann, and C. Berlin, "Call for Papers: Human-centric production and logistics systems," *Int. J. Prod. Res.*, 2022, [Online]. Available: https://www.callforpapers.co.uk/human-factors-i50

[3] Wang Baicun, Peng Tao, Xi Vincent Wang, Thorsten Wuest, David Romero, and Lihui Wang, Eds., "Human-centric Smart Manufacturing: Trends, Issues and Challenges," *J. Manuf. Syst.*, 2021.

[4] "Industry 5.0: Towards more sustainable, resilient and human-centric industry." https://research-and-innovation.ec.europa.eu/news/all-research-and-innovation-news/industry-50-towards-more-sustainable-resilient-and-human-centric-industry-2021-01-07_en (accessed Sep. 20, 2022).

[5] S. Folkard and D. A. Lombardi, "Modeling the impact of the components of long work hours on injuries and 'accidents,'" *Am. J. Ind. Med.*, vol. 49, no. 11, pp. 953–963, 2006, doi: 10.1002/ajim.20307.

[6] D. Fischer, D. A. Lombardi, S. Folkard, J. Willetts, and D. C. Christiani, "Updating the 'Risk Index': A systematic review and meta-analysis of occupational injuries and work schedule characteristics," *Chronobiol. Int.*, vol. 34, no. 10, pp. 1423–1438, 2017, doi: 10.1080/07420528.2017.1367305.

[7] T. Åkkerstedt, S. Folkard, and C. Portin, "Predictions from the Three-Process Model of Alertness," *Aviat. Space Environ. Med.*, vol. 75, no. 3, 2004.

[8] M. Y. Jaber, Z. S. Givi, and W. P. Neumann, "Incorporating human fatigue and recovery into the learning–forgetting process," *Appl. Math. Model.*, vol. 37, no. 12–13, pp. 7287–7299, Jul. 2013, doi: 10.1016/j.apm.2013.02.028.

[9] F. Fruggiero, S. Riemma, Y. Ouazene, R. Macchiaroli, and V. Guglielmi, "Incorporating the Human Factor within Manufacturing Dynamics," *IFAC-Pap.*, vol. 49, no. 12, pp. 1691–1696, 2016, doi: 10.1016/j.ifacol.2016.07.825.

[10] Z. Sedighi Maman, M. A. Alamdar Yazdi, L. A. Cavuoto, and F. M. Megahed, "A data-driven approach to modeling physical fatigue in the workplace using wearable sensors," *Appl. Ergon.*, vol. 65, pp. 515–529, 2017, doi: 10.1016/j.apergo.2017.02.001.

[11] E. Bal, O. Arslan, and L. Tavacioglu, "Prioritization of the causal factors of fatigue in seafarers and measurement of fatigue with the application of the Lactate Test," *Saf. Sci.*, vol. 72, pp. 46–54, 2015.

[12] M. Pal, R. Roy, J. Basu, and M. S. Bepari, "Blind source separation: A review and analysis," *2013 Int. Conf. Orient. COCOSDA Held Jointly 2013 Conf. Asian Spok. Lang. Res. Eval. O-COCOSDACASLRE 2013*, 2013, doi: 10.1109/ICSDA.2013.6709849.

[13] J. He, W. Chen, and Y. Song, "Single Channel Blind Source Separation Under Deep Recurrent Neural Network," *Wirel. Pers. Commun.*, vol. 115, no. 2, pp. 1277–1289, Nov. 2020, doi: 10.1007/S11277-020-07624-4/FIGURES/6.

[14] M. Cogswell, F. Ahmed, R. Girshick, L. Zitnick, and D. Batra, "Reducing overfitting in deep networks by decorrelating representations".

[15] S. Bianco, L. Celona, and P. Napoletano, "Disentangling Image distortions in deep feature space," *Pattern Recognit. Lett.*, vol. 148, pp. 128–135, Aug. 2021, doi: 10.1016/J.PATREC.2021.05.008.

[16] G. Lample, N. Zeghidour, N. Usunier, A. Bordes, L. Denoyer, and M. ' A. Ranzato, "Fader Networks: Manipulating Images by Sliding Attributes," in *Advances in Neural Information Processing Systems 30 (NIPS)*, 2017. Accessed: Dec. 27, 2022. [Online]. Available: https://github.com/facebookresearch/FaderNetworks

[17] Y. Liu, M. De Nadai, J. Yao, N. Sebe, B. Lepri, and X. Alameda-Pineda, "GMM-UNIT: Unsupervised Multi-Domain and Multi-Modal Image-to-Image Translation via Attribute Gaussian Mixture Modeling".

[18] T. Karras NVIDIA and S. Laine NVIDIA, "#StyleGAN - A Style-Based Generator Architecture for Generative Adversarial Networks Timo Aila NVIDIA," *Cvpr 2019*, 2019, [Online]. Available: https://github.com/NVlabs/stylegan

[19] I. Higgins *et al.*, "β-VAE: LEARNING BASIC VISUAL CONCEPTS WITH A CONSTRAINED VARIATIONAL FRAMEWORK," in *International conference on learning representations, ICLR*, 2017. Accessed: Dec. 28, 2022. [Online]. Available: https://openreview.net/forum?id=Sy2fzU9gl

[20] B. Liu, Y. Zhu, Z. Fu, G. De Melo, and A. Elgammal, "Disentangling GAN with One-Hot Sampling and Orthogonal Regularization", Accessed: Dec. 27, 2022. [Online]. Available: www.aaai.org

[21] H. Kim and A. Mnih, "Disentangling by Factorising," in *NIPS, Learning Disentangled Representations: From Perception to Control Workshop*, 2017.

[22] Y. F. Zhou, R. H. Jiang, X. Wu, J. Y. He, S. Weng, and Q. Peng, "BranchGAN: Unsupervised Mutual Image-to-Image Transfer with A Single Encoder and Dual Decoders," *IEEE Trans. Multimed.*, vol. 21, no. 12, pp. 3136–3149, Dec. 2019, doi: 10.1109/TMM.2019.2920613.

[23] M. Y. Liu, T. Breuel, and J. Kautz, "Unsupervised image-to-image translation networks," *Adv. Neural Inf. Process. Syst.*, vol. 2017-Decem, no. Nips, pp. 701–709, 2017.

[24] R. G. M. Morris, “D.O. Hebb: The Organization of Behavior, Wiley: New York; 1949,” *Brain Res. Bull.*, vol. 50, no. 5–6, p. 437, Nov. 1999, doi: 10.1016/S0361-9230(99)00182-3.

[25] E. Oja, “Simplified neuron model as a principal component analyzer,” *J. Math. Biol.*, vol. 15, no. 3, pp. 267–273, Nov. 1982, doi: 10.1007/BF00275687.

[26] A. Carlson, “Biological Cybernetics Anti-Hebbian learning in a nonlinear neural network,” 1990.

[27] C. Pehlevan, S. Mohan, and D. B. Chklovskii, “Blind Nonnegative Source Separation Using Biological Neural Networks,” *Neural Comput.*, vol. 29, no. 11, pp. 2925–2954, Nov. 2017, doi: 10.1162/neco_a_01007.

[28] B. Cheung, J. A. Livezey, A. K. Bansal, and B. A. Olshausen, “Discovering Hidden Factors of Variation in Deep Networks”.

[29] G. Alain, Y. Bengio, A. Courville, R. Fergus, and C. Manning, “What Regularized Auto-Encoders Learn from the Data-Generating Distribution,” *J. Mach. Learn. Res.*, vol. 15, pp. 3743–3773, 2014.

[30] R. E. Kalman, “A New Approach to Linear Filtering and Prediction Problems,” *J. Basic Eng.*, vol. 82, no. 1, p. 35, 1960, doi: 10.1115/1.3662552.

[31] B. A. McElhoe, “An assessment of the navigation and course corrections for a manned flyby of mars or venus,” *IEEE Trans. Aerosp. Electron. Syst.*, vol. AES-2, no. 4, pp. 613–623, 1966, doi: 10.1109/TAES.1966.4501892.

[32] P Del Moral, “nonlinear Filtering: Interacting Particle Resolution,” *Markov Process. Relat. Fields*, vol. 2, no. 4, pp. 555–580, 1996.

[33] G. Revach, N. Shlezinger, X. Ni, A. L. Escoriza, R. J. G. van Sloun, and Y. C. Eldar, “KalmanNet: Neural Network Aided Kalman Filtering for Partially Known Dynamics,” *IEEE Trans. Signal Process.*, vol. 70, pp. 1532–1547, 2022, doi: 10.1109/TSP.2022.3158588.

[34] K. Hassan, *nonlinear Systems*. Prentice-Hall, 2002.