

References

- [1] C. M. Bishop, *Pattern Recognition and Machine Learning*. New York: Springer, 2006.
- [2] Z. L. Tan and M. W. Mak, "Bottleneck features from SNR-adaptive denoising deep classifier for speaker identification," in *Proceedings of Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*, 2015.
- [3] Z. Tan, M. Mak, B. K. Mak, and Y. Zhu, "Denoised senone i-vectors for robust speaker verification," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 26, no. 4, pp. 820–830, Apr. 2018.
- [4] L. v. d. Maaten and G. Hinton, "Visualizing data using *t*-SNE," *Journal of Machine Learning Research*, vol. 9, pp. 2579–2605, Nov. 2008.
- [5] M. H. Moattar and M. M. Homayounpour, "A review on speaker diarization systems and approaches," *Speech Communication*, vol. 54, no. 10, pp. 1065–1103, 2012.
- [6] S. B. Davis and P. Mermelstein, "Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 28, no. 4, pp. 357–366, Aug. 1980.
- [7] D. A. Reynolds, T. F. Quatieri, and R. B. Dunn, "Speaker verification using adapted Gaussian mixture models," *Digital Signal Processing*, vol. 10, no. 1–3, pp. 19–41, Jan. 2000.
- [8] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 39, no. 1, pp. 1–38, 1977.
- [9] J. Pelecanos and S. Sridharan, "Feature warping for robust speaker verification," in *Proceedings of Speaker and Language Recognition Workshop (Odyssey)*, 2001, pp. 213–218.
- [10] M. W. Mak, K. K. Yiu, and S. Y. Kung, "Probabilistic feature-based transformation for speaker verification over telephone networks," *Neurocomputing: Special Issue on Neural Networks for Speech and Audio Processing*, vol. 71, pp. 137–146, 2007.
- [11] R. Teunen, B. Shahshahani, and L. Heck, "A model-based transformational approach to robust speaker recognition," in *Proc of International Conference on Spoken Language Processing (ICSLP)*, vol. 2, 2000, pp. 495–498.
- [12] K. K. Yiu, M. W. Mak, and S. Y. Kung, "Environment adaptation for robust speaker verification by cascading maximum likelihood linear regression and reinforced learning," *Computer Speech and Language*, vol. 21, pp. 231–246, 2007.
- [13] R. Auckenthaler, M. Carey, and H. Lloyd-Thomas, "Score normalization for text-independent speaker verification systems," *Digital Signal Processing*, vol. 10, no. 1–3, pp. 42–54, Jan. 2000.
- [14] W. M. Campbell, D. E. Sturim, and D. A. Reynolds, "Support vector machines using GMM supervectors for speaker verification," *IEEE Signal Processing Letters*, vol. 13, no. 5, pp. 308–311, May 2006.

- [15] P. Kenny, G. Boulianne, P. Ouellet, and P. Dumouchel, "Joint factor analysis versus eigen-channels in speaker recognition," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 15, no. 4, pp. 1435–1447, May 2007.
- [16] N. Dehak, P. Kenny, R. Dehak, P. Dumouchel, and P. Ouellet, "Front-end factor analysis for speaker verification," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 19, no. 4, pp. 788–798, May 2011.
- [17] S. Prince and J. Elder, "Probabilistic linear discriminant analysis for inferences about identity," in *Proceedings of IEEE International Conference on Computer Vision (ICCV)*, 2007, pp. 1–8.
- [18] A. Martin, G. Doddington, T. Kamm, M. Ordowski, and M. Przybocki, "The DET curve in assessment of detection task performance," in *Proceedings of European Conference on Speech Communication and Technology (EUROSPEECH)*, 1997, pp. 1895–1898.
- [19] D. Leeuwen and N. Brümmer, "The distribution of calibrated likelihood-ratios in speaker recognition," in *Proceedings of Annual Conference of International Speech Communication Association (INTERSPEECH)*, 2013, pp. 1619–1623.
- [20] K. Hornik, M. Stinchcombe, and H. White, "Multilayer feedforward networks are universal approximators," *Neural Networks*, vol. 2, pp. 359–366, 1989.
- [21] S. Kullback and R. A. Leibler, "On information and sufficiency," *Annals of Mathematical Statistics*, vol. 22, no. 1, pp. 79–86, 1951.
- [22] M. Jordan, Z. Ghahramani, T. Jaakkola, and L. Saul, "An introduction to variational methods for graphical models," *Machine Learning*, vol. 37, no. 2, pp. 183–233, 1999.
- [23] H. Attias, "Inferring parameters and structure of latent variable models by variational Bayes," in *Proceedings of Conference on Uncertainty in Artificial Intelligence (UAI)*, 1999, pp. 21–30.
- [24] R. M. Neal, "Probabilistic inference using Markov chain Monte Carlo methods," Department of Computer Science, University of Toronto, Tech. Rep., 1993.
- [25] J. S. Liu, *Monte Carlo Strategies in Scientific Computing*. New York, NY: Springer, 2008.
- [26] C. Andrieu, N. De Freitas, A. Doucet, and M. I. Jordan, "An introduction to MCMC for machine learning," *Machine Learning*, vol. 50, no. 1-2, pp. 5–43, 2003.
- [27] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 6, no. 1, pp. 721–741, 1984.
- [28] W. K. Hastings, "Monte Carlo sampling methods using Markov chains and their applications," *Biometrika*, vol. 57, pp. 97–109, 1970.
- [29] Y. W. Teh, M. I. Jordan, M. J. Beal, and D. M. Blei, "Hierarchical Dirichlet processes," *Journal of American Statistical Association*, vol. 101, no. 476, pp. 1566–1581, 2006.
- [30] S. Watanabe and J.-T. Chien, *Bayesian Speech and Language Processing*. Cambridge, UK: Cambridge University Press, 2015.
- [31] D. J. MacKay, "Bayesian interpolation," *Neural computation*, vol. 4, no. 3, pp. 415–447, 1992.
- [32] S. Y. Kung, M. W. Mak, and S. H. Lin, *Biometric Authentication: A Machine Learning Approach*. Englewood Cliffs, NJ: Prentice Hall, 2005.
- [33] V. N. Vapnik, *The Nature of Statistical Learning Theory*. New York: Springer-Verlag, 1995.
- [34] S. P. Boyd and L. Vandenberghe, *Convex Optimization*. New York: Cambridge University Press, 2004.
- [35] M. Mak and W. Rao, "Utterance partitioning with acoustic vector resampling for GMM-SVM speaker verification," *Speech Communication*, vol. 53, no. 1, pp. 119–130, Jan. 2011.

- [36] G. Wu and E. Y. Chang, "KBA: Kernel boundary alignment considering imbalanced data distribution," *IEEE Transactions on Knowledge and Data Engineering*, vol. 17, no. 6, pp. 786–795, 2005.
- [37] Y. Tang, Y. Q. Zhang, N. V. Chawla, and S. Krasser, "SVMs modeling for highly imbalanced classification," *IEEE Transactions on System, Man, and Cybernetics, Part B*, vol. 39, no. 1, pp. 281–288, Feb. 2009.
- [38] M. W. Mak and W. Rao, "Acoustic vector resampling for GMMSVM-based speaker verification," in *Proceedings of Annual Conference of International Speech Communication Association (INTERSPEECH)*, 2010, pp. 1449–1452.
- [39] W. Rao and M. W. Mak, "Addressing the data-imbalance problem in kernel-based speaker verification via utterance partitioning and speaker comparison," in *Interspeech*, 2011, pp. 2717–2720.
- [40] A. Solomonoff, C. Quillen, and W. M. Campbell, "Channel compensation for SVM speaker recognition," in *Proceedings of Speaker and Language Recognition Workshop (Odyssey)*, 2004, pp. 57–62.
- [41] A. Solomonoff, W. M. Campbell, and I. Boardman, "Advances in channel compensation for SVM speaker recognition," in *Proceedings of International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 2005, pp. 629–632.
- [42] W. M. Campbell, D. E. Sturim, D. A. Reynolds, and A. Solomonoff, "SVM based speaker verification using a GMM supervector kernel and NAP variability compensation," in *Proceedings of International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, vol. 1, 2006, pp. 97–100.
- [43] E. Kokiopoulou, J. Chen, and Y. Saad, "Trace optimization and eigenproblems in dimension reduction methods," *Numerical Linear Algebra with Applications*, vol. 18, no. 3, pp. 565–602, 2011.
- [44] P. Bromiley, "Products and convolutions of Gaussian probability density functions," *Tina-Vision Memo*, vol. 3, no. 4, 2003.
- [45] P. Kenny, G. Boulianne, and P. Dumouchel, "Eigenvoice modeling with sparse training data," *IEEE Transactions on Speech and Audio Processing*, vol. 13, no. 3, pp. 345–354, 2005.
- [46] S. M. Kay, *Fundamentals of Statistical Signal Processing*. Englewood Cliffs, NJ: Prentice-Hall, 1993.
- [47] M. W. Mak and J. T. Chien, "PLDA and mixture of PLDA formulations," Supplementary Materials for "Mixture of PLDA for Noise Robust I-Vector Speaker Verification," *IEEE/ACM Trans. on Audio Speech and Language Processing*, vol. 24, No. 1, pp. 130–142, Jan. 2016. [Online]. Available: <http://bioinfo.eie.polyu.edu.hk/mPLDA/SuppMaterials.pdf>
- [48] P. Rajan, A. Afanasyev, V. Hautamäki, and T. Kinnunen, "From single to multiple enrollment i-vectors: Practical PLDA scoring variants for speaker verification," *Digital Signal Processing*, vol. 31, pp. 93–101, 2014.
- [49] L. Chen, K. A. Lee, B. Ma, W. Guo, H. Li, and L. R. Dai, "Minimum divergence estimation of speaker prior in multi-session PLDA scoring," in *Proceedings of International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 2014, pp. 4007–4011.
- [50] S. Cumani, O. Plchot, and P. Laface, "On the use of i-vector posterior distributions in probabilistic linear discriminant analysis," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 22, no. 4, pp. 846–857, 2014.
- [51] L. Burget, O. Plchot, S. Cumani, O. Glembek, P. Matejka, and N. Brimmer, "Discriminatively trained probabilistic linear discriminant analysis for speaker verification," in

- Acoustics, Speech, and Signal Processing (ICASSP), 2011 IEEE International Conference on*, 2011, pp. 4832–4835.
- [52] V. Vasilakakis, P. Laface, and S. Cumani, “Pairwise discriminative speaker verification in the I-vector space,” *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 21, no. 6, pp. 1217–1227, 2013.
 - [53] J. Rohdin, S. Biswas, and K. Shinoda, “Constrained discriminative plda training for speaker verification,” in *Proceedings IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*. IEEE, 2014, pp. 1670–1674.
 - [54] N. Li and M. W. Mak, “SNR-invariant PLDA modeling for robust speaker verification,” in *Proceedings of Annual Conference of International Speech Communication Association (INTERSPEECH)*, 2015.
 - [55] N. Li and M. W. Mak, “SNR-invariant PLDA modeling in nonparametric subspace for robust speaker verification,” *IEEE/ACM Trans. on Audio Speech and Language Processing*, vol. 23, no. 10, pp. 1648–1659, 2015.
 - [56] S. O. Sadjadi, J. Pelecanos, and W. Zhu, “Nearest neighbor discriminant analysis for robust speaker recognition,” in *Proceedings of Annual Conference of International Speech Communication Association (INTERSPEECH)*, 2014, pp. 1860–1864.
 - [57] L. He, X. Chen, C. Xu, and J. Liu, “Multi-objective optimization training of plda for speaker verification,” in *2019 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*. IEEE, 2019, pp. 6026–6030.
 - [58] O. Ghahabi and J. Hernando, “Deep belief networks for i-vector based speaker recognition,” in *Proceedings of International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 2014, pp. 1700–1704.
 - [59] T. Stafylakis, P. Kenny, M. Senoussaoui, and P. Dumouchel, “Preliminary investigation of Boltzmann machine classifiers for speaker recognition,” in *Proceedings of Speaker and Language Recognition Workshop (Odyssey)*, 2012.
 - [60] O. Ghahabi and J. Hernando, “I-vector modeling with deep belief networks for multi-session speaker recognition,” in *Proceedings of Speaker and Language Recognition Workshop (Odyssey)*, 2014, pp. 305–310.
 - [61] P. Kenny, “Bayesian speaker verification with heavy-tailed priors,” in *Proceedings of Speaker and Language Recognition Workshop (Odyssey)*, 2010.
 - [62] N. Brummer, A. Silnova, L. Burget, and T. Stafylakis, “Gaussian meta-embeddings for efficient scoring of a heavy-tailed PLDA model,” in *Proceedings of Speaker and Language Recognition Workshop (Odyssey)*, 2018, pp. 349–356.
 - [63] K. B. Petersen and M. S. Pedersen, “The matrix cookbook,” Oct 2008. [Online]. Available: www2.imm.dtu.dk/pubdb/p.php?3274
 - [64] W. D. Penny, “KL-Divergences of Normal, Gamma, Dirichlet and Wishart densities,” Department of Cognitive Neurology, University College London, Tech. Rep., 2001.
 - [65] J. Soch and C. Alfeld, “Kullback-Leibler divergence for the normal-Gamma distribution,” *arXiv preprint arXiv:1611.01437*, 2016.
 - [66] D. Garcia-Romero and C. Espy-Wilson, “Analysis of i-vector length normalization in speaker recognition systems,” in *Proceedings of Annual Conference of International Speech Communication Association (INTERSPEECH)*, 2011, pp. 249–252.
 - [67] A. Silnova, N. Brummer, D. Garcia-Romero, D. Snyder, and L. Burget, “Fast variational Bayes for heavy-tailed PLDA applied to i-vectors and x-vectors,” *arXiv preprint arXiv:1803.09153*, 2018.

- [68] S. Shum, N. Dehak, E. Chuangsuwanich, D. Reynolds, and J. Glass, "Exploiting intra-conversation variability for speaker diarization," in *Proceedings of Annual Conference of International Speech Communication Association (INTERSPEECH)*, 2011, pp. 945–948.
- [69] E. Khoury and M. Garland, "I-vectors for speech activity detection," in *Proceedings of Speaker and Language Recognition Workshop (Odyssey)*, 2016, pp. 334–339.
- [70] N. Dehak, P. A. Torres-Carrasquillo, D. Reynolds, and R. Dehak, "Language recognition via i-vectors and dimensionality reduction," in *Proceedings of Annual Conference of International Speech Communication Association (INTERSPEECH)*, 2011, pp. 857–860.
- [71] S. S. Xu, M.-W. Mak, and C.-C. Cheung, "Patient-specific heartbeat classification based on i-vector adapted deep neural networks," in *Proceedings of IEEE International Conference on Bioinformatics and Biomedicine*, 2018.
- [72] P. Kenny, "A small footprint i-vector extractor," in *Proceedings of Speaker and Language Recognition Workshop (Odyssey)*, 2012.
- [73] J. Luttinen and A. Ilin, "Transformations in variational bayesian factor analysis to speed up learning," *Neurocomputing*, vol. 73, no. 7–9, pp. 1093–1102, 2010.
- [74] A. Hatch, S. Kajarekar, and A. Stolcke, "Within-class covariance normalization for SVM-based speaker recognition," in *Proceedings of International Conference on Spoken Language Processing (ICSLP)*, 2006, pp. 1471–1474.
- [75] K. Fukunaga, *Introduction to Statistical Pattern Recognition*. Boston, MA: Academic Press, 1990.
- [76] Z. Li, D. Lin, and X. Tang, "Nonparametric discriminant analysis for face recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 4, pp. 755–761, 2009.
- [77] F. Bahmaninezhad and J. H. Hansen, "I-vector/PLDA speaker recognition using support vectors with discriminant analysis," in *Proceedings of International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*. IEEE, 2017, pp. 5410–5414.
- [78] M. W. Mak and H. B. Yu, "A study of voice activity detection techniques for NIST speaker recognition evaluations," *Computer, Speech and Language*, vol. 28, no. 1, pp. 295–313, Jan. 2014.
- [79] W. Rao and M. W. Mak, "Boosting the performance of i-vector based speaker verification via utterance partitioning," *IEEE Trans. on Audio, Speech, and Language Processing*, vol. 21, no. 5, pp. 1012–1022, May 2013.
- [80] P. Kenny, T. Stafylakis, P. Ouellet, M. J. Alam, and P. Dumouchel, "PLDA for speaker verification with utterances of arbitrary duration," in *Proceedings of International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 2013, pp. 7649–7653.
- [81] W. Rao, M. W. Mak, and K. A. Lee, "Normalization of total variability matrix for i-vector/PLDA speaker verification," in *Proceedings of International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 2015, pp. 4180–4184.
- [82] W. W. Lin and M. W. Mak, "Fast scoring for PLDA with uncertainty propagation," in *Proceedings of Speaker and Language Recognition Workshop (Odyssey)*, 2016, pp. 31–38.
- [83] W. W. Lin, M. W. Mak, and J. T. Chien, "Fast scoring for PLDA with uncertainty propagation via i-vector grouping," *Computer Speech & Language*, vol. 45, pp. 503–515, 2017.
- [84] Y. Lei, N. Scheffer, L. Ferrer, and M. McLaren, "A novel scheme for speaker recognition using a phonetically-aware deep neural network," in *Proceedings of International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 2014.

- [85] L. Ferrer, Y. Lei, M. McLaren, and N. Scheffer, "Study of senone-based deep neural network approaches for spoken language recognition," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 24, no. 1, pp. 105–116, 2016.
- [86] P. Kenny, "Joint factor analysis of speaker and session variability: Theory and algorithms," CRIM, Montreal, Tech. Rep. CRIM-06/08-13, 2005.
- [87] P. Kenny, P. Ouellet, N. Dehak, V. Gupta, and P. Dumouchel, "A study of inter-speaker variability in speaker verification," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 16, no. 5, pp. 980–988, 2008.
- [88] O. Glembek, L. Burget, N. Dehak, N. Brummer, and P. Kenny, "Comparison of scoring methods used in speaker recognition with joint factor analysis," in *Proceedings of International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 2009, pp. 4057–4060.
- [89] G. E. Hinton and R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *Science*, vol. 313, no. 5786, pp. 504–507, 2006.
- [90] G. E. Hinton, *A Practical Guide to Training Restricted Boltzmann Machines*. Berlin Heidelberg: Springer, 2012, pp. 599–619.
- [91] J. J. Hopfield, "Neural networks and physical systems with emergent collective computational abilities," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 79, no. 8, pp. 2554–2558, 1982.
- [92] G. E. Hinton, "Training products of experts by minimizing contrastive divergence," *Neural Computation*, vol. 14, no. 8, pp. 1771–1800, 2002.
- [93] M. A. Carreira-Perpinan and G. E. Hinton, "On contrastive divergence learning," in *Proceedings of International Workshop on Artificial Intelligence and Statistics (AISTATS)*, 2005, pp. 33–40.
- [94] G. E. Hinton, S. Osindero, and Y.-W. Teh, "A fast learning algorithm for deep belief nets," *Neural Computation*, vol. 18, no. 7, pp. 1527–1554, 2006.
- [95] N. Li, M. W. Mak, and J. T. Chien, "DNN-driven mixture of PLDA for robust speaker verification," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, no. 6, pp. 1371–1383, 2017.
- [96] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, pp. 2278–2324, 1998.
- [97] D. Yu, G. Hinton, N. Morgan, J.-T. Chien, and S. Sagayama, "Introduction to the special section on deep learning for speech and language processing," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 20, no. 1, pp. 4–6, 2012.
- [98] G. Hinton, L. Deng, D. Yu, G. Dahl, A. Mohamed, N. Jaitly, A. Senior, V. Vanhoucke, P. Nguyen, T. Sainath, and B. Kingsbury, "Deep neural networks for acoustic modeling in speech recognition: Four research groups share their views," *IEEE Signal Processing Magazine*, vol. 29, no. 6, pp. 82–97, 2012.
- [99] G. Saon and J.-T. Chien, "Large-vocabulary continuous speech recognition systems: A look at some recent advances," *IEEE Signal Processing Magazine*, vol. 29, no. 6, pp. 18–33, 2012.
- [100] J.-T. Chien and Y.-C. Ku, "Bayesian recurrent neural network for language modeling," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 27, no. 2, pp. 361–374, 2016.
- [101] M. D. Zeiler, G. W. Taylor, and R. Fergus, "Adaptive deconvolutional networks for mid and high level feature learning," in *Proceedings of IEEE International Conference on Computer Vision (ICCV)*, 2011, pp. 2018–2025.

- [102] J. Xie, L. Xu, and E. Chen, "Image denoising and inpainting with deep neural networks," in *Advances in Neural Information Processing Systems 25*, F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, Eds., 2012, pp. 341–349.
- [103] R. Salakhutdinov and H. Larochelle, "Efficient learning of deep Boltzmann machines," in *Proceedings of International Conference on Artificial Intelligence and Statistics (AISTATS)*, 2010, pp. 693–700.
- [104] P. Vincent, H. Larochelle, I. Lajoie, Y. Bengio, and P.-A. Manzagol, "Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion," *Journal of Machine Learning Research*, vol. 11, pp. 3371–3408, 2010.
- [105] M. Schuster and K. K. Paliwal, "Bidirectional recurrent neural networks," *IEEE Transactions on Signal Processing*, vol. 45, no. 11, pp. 2673–2681, 1997.
- [106] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning internal representation by backpropagating errors," *Nature*, vol. 323, pp. 533–536, 1986.
- [107] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, MA: MIT Press, 2016.
- [108] Y. Bengio, P. Lamblin, D. Popovici, and H. Larochelle, "Greedy layer-wise training of deep networks," in *Advances in Neural Information Processing Systems 19*, B. Schölkopf, J. C. Platt, and T. Hoffman, Eds. Cambridge, MA: MIT Press, 2007, pp. 153–160.
- [109] G. E. Hinton and R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *Science*, vol. 313, pp. 504–507, 2006.
- [110] G. E. Hinton, S. Osindero, and Y.-W. Teh, "A fast learning algorithm for deep belief nets," *Neural Computation*, vol. 18, no. 7, pp. 1527–1554, 2006.
- [111] R. Salakhutdinov and G. E. Hinton, "Deep Boltzmann machines," in *Proceedings of International Conference on Artificial Intelligence and Statistics (AISTATS)*, 2009, p. 3.
- [112] P. Vincent, H. Larochelle, Y. Bengio, and P.-A. Manzagol, "Extracting and composing robust features with denoising autoencoders," in *Proceedings of International Conference on Machine Learning (ICML)*, 2008, pp. 1096–1103.
- [113] A. Hyvärinen, "Estimation of non-normalized statistical models by score matching," *Journal of Machine Learning Research*, vol. 6, pp. 695–709, 2005.
- [114] D. P. Kingma and M. Welling, "Auto-encoding variational Bayes," in *Proceedings of International Conference on Learning Representation (ICLR)*, 2014.
- [115] J.-T. Chien and K.-T. Kuo, "Variational recurrent neural networks for speech separation," in *Proceedings of Annual Conference of International Speech Communication Association (INTERSPEECH)*, 2017, pp. 1193–1197.
- [116] J.-T. Chien and C.-W. Hsu, "Variational manifold learning for speaker recognition," in *Proceedings of International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 2017, pp. 4935–4939.
- [117] D. J. Rezende, S. Mohamed, and D. Wierstra, "Stochastic backpropagation and approximate inference in deep generative models," in *Proceedings of International Conference on Machine Learning (ICML)*, 2014, pp. 1278–1286.
- [118] 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)*, 2014, pp. 2672–2680.
- [119] J.-T. Chien and K.-T. Peng, "Adversarial manifold learning for speaker recognition," in *Prof. of IEEE Automatic Speech Recognition and Understanding Workshop (ASRU)*, 2017, pp. 599–605.

- [120] J.-T. Chien and K.-T. Peng, "Adversarial learning and augmentation for speaker recognition," in *Proceedings of Speaker and Language Recognition Workshop (Odyssey)*, 2018, pp. 342–348.
- [121] Y. Bengio, E. Laufer, G. Alain, and J. Yosinski, "Deep generative stochastic networks trainable by backprop," in *Proceedings of International Conference on Machine Learning (ICML)*, 2014, pp. 226–234.
- [122] A. Makhzani, J. Shlens, N. Jaitly, and I. Goodfellow, "Adversarial autoencoders," *arXiv preprint arXiv:1511.05644*, 2015.
- [123] A. B. L. Larsen, S. K. Sønderby, and O. Winther, "Autoencoding beyond pixels using a learned similarity metric," in *Proceedings of International Conference on Machine Learning (ICML)*, no. 1558–1566, 2015.
- [124] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Transactions on Knowledge and Data Engineering*, vol. 22, no. 10, pp. 1345–1359, 2009.
- [125] A. Evgeniou and M. Pontil, "Multi-task feature learning," *Advances in Neural Information Processing Systems (NIPS)*, vol. 19, p. 41, 2007.
- [126] R. K. Ando and T. Zhang, "A framework for learning predictive structures from multiple tasks and unlabeled data," *Journal of Machine Learning Research*, vol. 6, pp. 1817–1853, 2005.
- [127] A. Argyriou, M. Pontil, Y. Ying, and C. A. Micchelli, "A spectral regularization framework for multi-task structure learning," in *Advances in Neural Information Processing Systems (NIPS)*, 2007, pp. 25–32.
- [128] W. Lin, M. Mak, and J. Chien, "Multisource i-vectors domain adaptation using maximum mean discrepancy based autoencoders," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 26, no. 12, pp. 2412–2422, Dec 2018.
- [129] W. W. Lin, M. W. Mak, L. X. Li, and J. T. Chien, "Reducing domain mismatch by maximum mean discrepancy based autoencoders," in *Proceedings of Speaker and Language Recognition Workshop (Odyssey)*, 2018, pp. 162–167.
- [130] M. Sugiyama, S. Nakajima, H. Kashima, P. V. Buenau, and M. Kawanabe, "Direct importance estimation with model selection and its application to covariate shift adaptation," in *Advances in Neural Information Processing Systems (NIPS)*, 2008, pp. 1433–1440.
- [131] S. Bickel, M. Brückner, and T. Scheffer, "Discriminative learning under covariate shift," *Journal of Machine Learning Research*, vol. 10, pp. 2137–2155, 2009.
- [132] J. Blitzer, R. McDonald, and F. Pereira, "Domain adaptation with structural correspondence learning," in *Proceedings of Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2006, pp. 120–128.
- [133] P. von Büna, F. C. Meinecke, F. C. Király, and K.-R. Müller, "Finding stationary subspaces in multivariate time series," *Physical Review Letters*, vol. 103, no. 21, p. 214101, 2009.
- [134] S. J. Pan, J. T. Kwok, and Q. Yang, "Transfer learning via dimensionality reduction," in *Proceedings of AAAI Conference on Artificial Intelligence*, vol. 8, 2008, pp. 677–682.
- [135] A. Gretton, K. M. Borgwardt, M. Rasch, B. Schölkopf, and A. J. Smola, "A kernel method for the two-sample-problem," in *Advances in Neural Information Processing Systems (NIPS)*, 2007, pp. 513–520.
- [136] K. M. Borgwardt, A. Gretton, M. J. Rasch, H.-P. Kriegel, B. Schölkopf, and A. J. Smola, "Integrating structured biological data by kernel maximum mean discrepancy," *Bioinformatics*, vol. 22, no. 14, pp. e49–e57, 2006.

- [137] A. Ahmed, K. Yu, W. Xu, Y. Gong, and E. Xing, "Training hierarchical feed-forward visual recognition models using transfer learning from pseudo-tasks," in *Proceedings of European Conference on Computer Vision (ECCV)*, 2008, pp. 69–82.
- [138] S. Ji, W. Xu, M. Yang, and K. Yu, "3D convolutional neural networks for human action recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 1, pp. 221–231, 2013.
- [139] G. Hinton, L. Deng, D. Yu, G. Dahl, A. Mohamed, N. Jaitly, A. Senior, V. Vanhoucke, P. Nguyen, T. Sainath, and B. Kingsbury, "Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups," *IEEE Signal Processing Magazine*, vol. 29, pp. 82–97, 2012.
- [140] G. E. Dahl, D. Yu, L. Deng, and A. Acero, "Context-dependent pre-trained deep neural networks for large-vocabulary speech recognition," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 20, no. 1, pp. 30–42, 2012.
- [141] L. Deng, "A tutorial survey of architectures, algorithms, and applications for deep learning," *APSIPA Transactions on Signal and Information Processing*, vol. 3, p. e2, 2014.
- [142] A. R. Mohamed, G. E. Dahl, and G. Hinton, "Acoustic modeling using deep belief networks," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 20, no. 1, pp. 14–22, 2012.
- [143] Y. Z. Işık, H. Erdogan, and R. Sarikaya, "S-vector: A discriminative representation derived from i-vector for speaker verification," in *Proceedings of European Signal Processing Conference (EUSIPCO)*, 2015, pp. 2097–2101.
- [144] S. Novoselov, T. Pekhovsky, O. Kudashev, V. S. Mendelev, and A. Prudnikov, "Non-linear PLDA for i-vector speaker verification," in *Proceedings of Annual Conference of the International Speech Communication Association (INTERSPEECH)*, 2015.
- [145] T. Pekhovsky, S. Novoselov, A. Sholohov, and O. Kudashev, "On autoencoders in the i-vector space for speaker recognition," in *Proceedings of Speaker and Language Recognition Workshop (Odyssey)*, 2016, pp. 217–224.
- [146] S. Mahto, H. Yamamoto, and T. Koshinaka, "I-vector transformation using a novel discriminative denoising autoencoder for noise-robust speaker recognition," in *Proceedings of Annual Conference of International Speech Communication Association (INTERSPEECH)*, 2017, pp. 3722–3726.
- [147] Y. Tian, M. Cai, L. He, and J. Liu, "Investigation of bottleneck features and multilingual deep neural networks for speaker verification," in *Proceedings of Annual Conference of International Speech Communication Association (INTERSPEECH)*, 2015, pp. 1151–1155.
- [148] Z. L. Tan, Y. K. Zhu, M. W. Mak, and B. Mak, "Senone i-vectors for robust speaker verification," in *Proceedings of International Symposium on Chinese Spoken Language Processing (ISCSLP)*, Tianjin, China, October 2016.
- [149] S. Yaman, J. Pelecanos, and R. Sarikaya, "Bottleneck features for speaker recognition," in *Proceedings of Speaker and Language Recognition Workshop (Odyssey)*, vol. 12, 2012, pp. 105–108.
- [150] E. Variani, X. Lei, E. McDermott, I. Lopez M., and J. Gonzalez-Dominguez, "Deep neural networks for small footprint text-dependent speaker verification," in *Proceedings of International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 2014, pp. 4052–4056.
- [151] T. Yamada, L. B. Wang, and A. Kai, "Improvement of distant-talking speaker identification using bottleneck features of DNN," in *Proceedings of Annual Conference of International Speech Communication Association (INTERSPEECH)*, 2013, pp. 3661–3664.

- [152] P. Kenny, V. Gupta, T. Stafylakis, P. Ouellet, and J. Alam, "Deep neural networks for extracting Baum-Welch statistics for speaker recognition," in *Proceedings of Speaker and Language Recognition Workshop (Odyssey)*, 2014, pp. 293–298.
- [153] D. Garcia-Romero and A. McCree, "Insights into deep neural networks for speaker recognition," in *Proceedings of Annual Conference of International Speech Communication Association (INTERSPEECH)*, 2015, pp. 1141–1145.
- [154] M. McLaren, Y. Lei, and L. Ferrer, "Advances in deep neural network approaches to speaker recognition," in *Proceedings of International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 2015, pp. 4814–4818.
- [155] D. Snyder, D. Garcia-Romero, D. Povey, and S. Khudanpur, "Deep neural network embeddings for text-independent speaker verification," in *Proceedings of Annual Conference of International Speech Communication Association (INTERSPEECH)*, 2017, pp. 999–1003.
- [156] D. Snyder, D. Garcia-Romero, G. Sell, D. Povey, and S. Khudanpur, "X-vectors: Robust DNN embeddings for speaker recognition," in *Proceedings of International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 2018, pp. 5329–5333.
- [157] Y. Tang, G. Ding, J. Huang, X. He, and B. Zhou, "Deep speaker embedding learning with multi-level pooling for text-independent speaker verification," in *Proceedings IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 2019, pp. 6116–6120.
- [158] C.-P. Chen, S.-Y. Zhang, C.-T. Yeh, J.-C. Wang, T. Wang, and C.-L. Huang, "Speaker characterization using tdnn- lstm based speaker embedding," in *Proceedings IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*. IEEE, 2019, pp. 6211–6215.
- [159] W. Zhu and J. Pelecanos, "A bayesian attention neural network layer for speaker recognition," in *Proceedings IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*. IEEE, 2019, pp. 6241–6245.
- [160] Y. Zhu, T. Ko, D. Snyder, B. Mak, and D. Povey, "Self-attentive speaker embeddings for text-independent speaker verification," in *Proceedings Interspeech*, vol. 2018, 2018, pp. 3573–3577.
- [161] N. Brummer, L. Burget, P. Garcia, O. Plchot, J. Rohdin, D. Romero, D. Snyder, T. Stafylakis, A. Swart, and J. Villalba, "Meta-embeddings: A probabilistic generalization of embeddings in machine learning," in *JHU HLTCOE 2017 SCALE Workshop*, 2017.
- [162] N. Li and M. W. Mak, "SNR-invariant PLDA modeling for robust speaker verification," in *Proceedings of Annual Conference of International Speech Communication Association (INTERSPEECH)*, 2015, pp. 2317–2321.
- [163] N. Li, M. W. Mak, W. W. Lin, and J. T. Chien, "Discriminative subspace modeling of SNR and duration variabilities for robust speaker verification," *Computer Speech & Language*, vol. 45, pp. 83–103, 2017.
- [164] S. Prince and J. Elder, "Probabilistic linear discriminant analysis for inferences about identity," in *Proceedings of IEEE International Conference on Computer Vision (ICCV)*, 2007, pp. 1–8.
- [165] S. J. Prince, *Computer Vision: Models, Learning, and Inference*. New York: Cambridge University Press, 2012.
- [166] A. Sizov, K. A. Lee, and T. Kinnunen, "Unifying probabilistic linear discriminant analysis variants in biometric authentication," in *Structural, Syntactic, and Statistical Pattern Recognition*. Berlin, Heidelberg: Springer, 2014, pp. 464–475.

- [167] T. Hasan, R. Saeidi, J. H. L. Hansen, and D. A. van Leeuwen, "Duration mismatch compensation for I-vector based speaker recognition system," in *Proceedings of International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 2013, pp. 7663–7667.
- [168] A. Kanagasundaram, D. Dean, S. Sridharan, J. Gonzalez-Dominguez, J. Gonzalez-Rodriguez, and D. Ramos, "Improving short utterance i-vector speaker verification using utterance variance modelling and compensation techniques," *Speech Communication*, vol. 59, pp. 69–82, 2014.
- [169] K. H. Norwich, *Information, Sensation, and Perception*. San Diego: Academic Press, 1993.
- [170] P. Billingsley, *Probability and Measure*. New York: John Wiley & Sons, 2008.
- [171] M. W. Mak, X. M. Pang, and J. T. Chien, "Mixture of PLDA for noise robust i-vector speaker verification," *IEEE/ACM Trans. on Audio Speech and Language Processing*, vol. 24, no. 1, pp. 132–142, 2016.
- [172] M. W. Mak, "SNR-dependent mixture of PLDA for noise robust speaker verification," in *Proceedings of Annual Conference of International Speech Communication Association (INTERSPEECH)*, 2014, pp. 1855–1859.
- [173] X. M. Pang and M. W. Mak, "Noise robust speaker verification via the fusion of SNR-independent and SNR-dependent PLDA," *International Journal of Speech Technology*, vol. 18, no. 4, 2015.
- [174] M. E. Tipping and C. M. Bishop, "Mixtures of probabilistic principal component analyzers," *Neural Computation*, vol. 11, no. 2, pp. 443–482, 1999.
- [175] T. Pekhovsky and A. Sizov, "Comparison between supervised and unsupervised learning of probabilistic linear discriminant analysis mixture models for speaker verification," *Pattern Recognition Letters*, vol. 34, no. 11, pp. 1307–1313, 2013.
- [176] N. Li, M. W. Mak, and J. T. Chien, "Deep neural network driven mixture of PLDA for robust i-vector speaker verification," in *Proceedings of IEEE Workshop on Spoken Language Technology (SLT)*, San Diego, CA, 2016, pp. 186–191.
- [177] S. Cumani and P. Laface, "Large-scale training of pairwise support vector machines for speaker recognition," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 22, no. 11, pp. 1590–1600, 2014.
- [178] D. Snyder, P. Ghahremani, D. Povey, D. Garcia-Romero, Y. Carmiel, and S. Khudanpur, "Deep neural network-based speaker embeddings for end-to-end speaker verification," in *Proceedings of IEEE Spoken Language Technology Workshop (SLT)*, 2016, pp. 165–170.
- [179] M. I. Mandasari, R. Saeidi, M. McLaren, and D. A. van Leeuwen, "Quality measure functions for calibration of speaker recognition systems in various duration conditions," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 21, no. 11, pp. 2425–2438, Nov. 2013.
- [180] M. I. Mandasari, R. Saeidi, and D. A. van Leeuwen, "Quality measures based calibration with duration and noise dependency for speaker recognition," *Speech Communication*, vol. 72, pp. 126–137, 2015.
- [181] A. O. J. Villalba, A. Miguel and E. Lleida, "Bayesian networks to model the variability of speaker verification scores in adverse environments," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 24, no. 12, pp. 2327–2340, 2016.
- [182] A. Nautsch, R. Saeidi, C. Rathgeb, and C. Busch, "Robustness of quality-based score calibration of speaker recognition systems with respect to low-SNR and short-duration conditions," in *Proceedings of Speaker and Language Recognition Workshop (Odyssey)*, 2016, pp. 358–365.

- [183] L. Ferrer, L. Burget, O. Plhot, and N. Scheffer, "A unified approach for audio characterization and its application to speaker recognition," in *Proceedings of Speaker and Language Recognition Workshop (Odyssey)*, 2012, pp. 317–323.
- [184] Q. Hong, L. Li, M. Li, L. Huang, L. Wan, and J. Zhang, "Modified-prior PLDA and score calibration for duration mismatch compensation in speaker recognition system," in *Proceedings of Annual Conference of International Speech Communication Association (INTERSPEECH)*, 2015.
- [185] A. Shulipa, S. Novoselov, and Y. Matveev, "Scores calibration in speaker recognition systems," in *Proceedings of International Conference on Speech and Computer*, 2016, pp. 596–603.
- [186] N. Brümmer, A. Swart, and D. van Leeuwen, "A comparison of linear and non-linear calibrations for speaker recognition," in *Proceedings of Speaker and Language Recognition Workshop (Odyssey)*, 2014, pp. 14–18.
- [187] N. Brümmer and G. Doddington, "Likelihood-ratio calibration using prior-weighted proper scoring rules," in *Proceedings of Annual Conference of International Speech Communication Association (INTERSPEECH)*, 2013, pp. 1976–1980.
- [188] N. Brümmer and D. Garcia-Romero, "Generative modelling for unsupervised score calibration," in *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 2014, pp. 1680–1684.
- [189] R. Caruana, "Multitask learning: A knowledge-based source of inductive bias," *Machine Learning*, vol. 28, pp. 41–75, 1997.
- [190] D. Chen and B. Mak, "Multitask learning of deep neural networks for low-resource speech recognition," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 23, no. 7, pp. 1172–1183, 2015.
- [191] Q. Yao and M. W. Mak, "SNR-invariant multitask deep neural networks for robust speaker verification," *IEEE Signal Processing Letters*, vol. 25, no. 11, pp. 1670–1674, Nov. 2018.
- [192] D. Garcia-Romero and A. McCree, "Supervised domain adaptation for i-vector based speaker recognition," in *Proceedings of International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 2014, pp. 4047–4051.
- [193] J. Villalba and E. Lleida, "Bayesian adaptation of PLDA based speaker recognition to domains with scarce development data," in *Proceedings of Speaker and Language Recognition Workshop (Odyssey)*, Singapore, 2012.
- [194] J. Villalba and E. Lleida, "Unsupervised adaptation of PLDA by using variational Bayes methods," in *Proceedings of International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 2014, pp. 744–748.
- [195] B. J. Borgström, E. Singer, D. Reynolds, and O. Sadjadi, "Improving the effectiveness of speaker verification domain adaptation with inadequate in-domain data," in *Proceedings of Annual Conference of International Speech Communication Association (INTERSPEECH)*, 2017, pp. 1557–1561.
- [196] S. Shum, D. A. Reynolds, D. Garcia-Romero, and A. McCree, "Unsupervised clustering approaches for domain adaptation in speaker recognition systems," in *Proceedings of Speaker and Language Recognition Workshop (Odyssey)*, 2014, pp. 266–272.
- [197] D. Garcia-Romero, X. Zhang, A. McCree, and D. Povey, "Improving speaker recognition performance in the domain adaptation challenge using deep neural networks," in *Proceedings of IEEE Spoken Language Technology Workshop (SLT)*. IEEE, 2014, pp. 378–383.

- [198] Q. Q. Wang and T. Koshinaka, “Unsupervised discriminative training of PLDA for domain adaptation in speaker verification,” in *Proceedings of Annual Conference of International Speech Communication Association (INTERSPEECH)*, 2017, pp. 3727–3731.
- [199] S. Shon, S. Mun, W. Kim, and H. Ko, “Autoencoder based domain adaptation for speaker recognition under insufficient channel information,” in *Proceedings of Annual Conference of International Speech Communication Association (INTERSPEECH)*, 2017, pp. 1014–1018.
- [200] H. Aronowitz, “Inter dataset variability compensation for speaker recognition,” in *Proceedings of International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 2014, pp. 4002–4006.
- [201] H. Aronowitz, “Compensating inter-dataset variability in PLDA hyper-parameters for robust speaker recognition,” in *Proceedings of Speaker and Language Recognition Workshop (Odyssey)*, 2014, pp. 282–286.
- [202] H. Rahman, A. Kanagasundaram, D. Dean, and S. Sridharan, “Dataset-invariant covariance normalization for out-domain PLDA speaker verification,” in *Proceedings of Annual Conference of International Speech Communication Association (INTERSPEECH)*, 2015, pp. 1017–1021.
- [203] A. Kanagasundaram, D. Dean, and S. Sridharan, “Improving out-domain PLDA speaker verification using unsupervised inter-dataset variability compensation approach,” in *Proceedings of International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 2015, pp. 4654–4658.
- [204] O. Glembek, J. Ma, P. Matejka, B. Zhang, O. Plchot, L. Burget, and S. Matsoukas, “Domain adaptation via within-class covariance correction in i-vector based speaker recognition systems,” in *Proceedings of International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 2014, pp. 4032–4036.
- [205] E. Singer and D. A. Reynolds, “Domain mismatch compensation for speaker recognition using a library of whiteners,” *IEEE Signal Processing Letters*, vol. 22, no. 11, pp. 2000–2003, 2015.
- [206] F. Bahmaninezhad and J. H. L. Hansen, “Compensation for domain mismatch in text-independent speaker recognition,” in *Proceedings of Annual Conference of International Speech Communication Association (INTERSPEECH)*, 2018, pp. 1071–1075.
- [207] H. Yu, Z. H. Tan, Z. Y. Ma, and J. Guo, “Adversarial network bottleneck features for noise robust speaker verification,” *arXiv preprint arXiv:1706.03397*, 2017.
- [208] D. Michelsanti and Z. H. Tan, “Conditional generative adversarial networks for speech enhancement and noise-robust speaker verification,” *arXiv preprint arXiv:1709.01703*, 2017.
- [209] J. C. Zhang, N. Inoue, and K. Shinoda, “I-vector transformation using conditional generative adversarial networks for short utterance speaker verification,” in *Proceedings Interspeech*, 2018, pp. 3613–3617.
- [210] Z. Meng, J. Y. Li, Z. Chen, Y. Zhao, V. Mazalov, Y. F. Gong, and B. H. Juang, “Speaker-invariant training via adversarial learning,” *arXiv preprint arXiv:1804.00732*, 2018.
- [211] Q. Wang, W. Rao, S. Sun, L. Xie, E. S. Chng, and H. Z. Li, “Unsupervised domain adaptation via domain adversarial training for speaker recognition,” in *Proceedings of International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 2018, pp. 4889–4893.
- [212] I. Viñals, A. Ortega, J. Villalba, A. Miguel, and E. Lleida, “Domain adaptation of PLDA models in broadcast diarization by means of unsupervised speaker clustering,” in

- Proceedings of Annual Conference of International Speech Communication Association (INTERSPEECH)*, 2017, pp. 2829–2833.
- [213] J. Y. Li, M. L. Seltzer, X. Wang, R. Zhao, and Y. F. Gong, “Large-scale domain adaptation via teacher-student learning,” in *Proceedings of Annual Conference of International Speech Communication Association (INTERSPEECH)*, 2017, pp. 2386–2390.
- [214] H. Aronowitz, “Inter dataset variability modeling for speaker recognition,” in *Proceedings International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 2017, pp. 5400–5404.
- [215] M. McLaren and D. Van Leeuwen, “Source-normalized LDA for robust speaker recognition using i-vectors from multiple speech sources,” *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 20, no. 3, pp. 755–766, 2012.
- [216] M. H. Rahman, I. Himawan, D. Dean, C. Fookes, and S. Sridharan, “Domain-invariant i-vector feature extraction for PLDA speaker verification,” in *Proceedings of Speaker and Language Recognition Workshop (Odyssey)*, 2018, pp. 155–161.
- [217] S. E. Shepstone, K. A. Lee, H. Li, Z.-H. Tan, and S. H. Jensen, “Total variability modeling using source-specific priors,” *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 24, no. 3, pp. 504–517, 2016.
- [218] M. J. Alam, G. Bhattacharya, and P. Kenny, “Speaker verification in mismatched conditions with frustratingly easy domain adaptation,” in *Proceedings of Speaker and Language Recognition Workshop (Odyssey)*, 2018, pp. 176–180.
- [219] B. Sun, J. Feng, and K. Saenko, “Return of frustratingly easy domain adaptation,” in *Proceedings of AAAI Conference on Artificial Intelligence*, vol. 6, no. 7, 2016.
- [220] J. Alam, P. Kenny, G. Bhattacharya, and M. Kockmann, “Speaker verification under adverse conditions using i-vector adaptation and neural networks,” in *Proceedings of Annual Conference of International Speech Communication Association (INTERSPEECH)*, 2017, pp. 3732–3736.
- [221] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. A. Alemi, “Inception-v4, inception-resnet and the impact of residual connections on learning,” in *Proceedings of AAAI Conference on Artificial Intelligence*, 2017.
- [222] G. Bhattacharya, J. Alam, P. Kenn, and V. Gupta, “Modelling speaker and channel variability using deep neural networks for robust speaker verification,” in *Proceedings of IEEE Spoken Language Technology Workshop (SLT)*, 2016, pp. 192–198.
- [223] *Domain Adaptation Challenge*, John Hopkins University, 2013.
- [224] A. Storkey, “When training and test sets are different: Characterizing learning transfer,” in *Dataset Shift in Machine Learning*, J. Quinonero-Candela, M. Sugiyama, A. Schwaighofer, and N. Lawrence, Eds. Cambridge, MA: MIT Press, 2009, pp. 3–28.
- [225] H. Shimodaira, “Improving predictive inference under covariate shift by weighting the log-likelihood function,” *Journal of Statistical Planning and Inference*, vol. 90, no. 2, pp. 227–244, 2000.
- [226] S. B. David, T. Lu, T. Luu, and D. Pál, “Impossibility theorems for domain adaptation,” in *Proceedings International Conference on Artificial Intelligence and Statistics (AISTATS)*, 2010, pp. 129–136.
- [227] Y. Mansour, M. Mohri, and A. Rostamizadeh, “Domain adaptation: Learning bounds and algorithms,” *arXiv preprint arXiv:0902.3430*, 2009.
- [228] P. Germain, A. Habrard, F. Laviolette, and E. Morvant, “A PAC-Bayesian approach for domain adaptation with specialization to linear classifiers,” in *Proceedings International Conference on Machine Learning (ICML)*, 2013, pp. 738–746.

- [229] H.-Y. Chen and J.-T. Chien, "Deep semi-supervised learning for domain adaptation," in *IEEE International Workshop on Machine Learning for Signal Processing (MLSP)*, 2015, pp. 1–6.
- [230] A. Gretton, K. M. Borgwardt, M. Rasch, B. Schölkopf, and A. J. Smola, "A kernel method for the two-sample-problem," in *Advances in Neural Information Processing Systems (NIPS)*, 2007, pp. 513–520.
- [231] Y. Li, K. Swersky, and R. Zemel, "Generative moment matching networks," in *Proceedings International Conference on Machine Learning (ICML)*, 2015, pp. 1718–1727.
- [232] M. Long, Y. Cao, J. Wang, and M. Jordan, "Learning transferable features with deep adaptation networks," in *Proceedings International Conference on Machine Learning (ICML)*, 2015, pp. 97–105.
- [233] A. Smola, A. Gretton, L. Song, and B. Schölkopf, "A Hilbert space embedding for distributions," in *International Conference on Algorithmic Learning Theory*. Berlin, Heidelberg: Springer, 2007, pp. 13–31.
- [234] P. Vincent, H. Larochelle, I. Lajoie, Y. Bengio, and P. Manzagol, "Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion," *Journal of Machine Learning Research*, vol. 11, pp. 3371–3408, 2010.
- [235] F. Schroff, D. Kalenichenko, and J. Philbin, "Facenet: A unified embedding for face recognition and clustering," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015, pp. 815–823.
- [236] Y. Wen, K. Zhang, Z. Li, and Y. Qiao, "A discriminative feature learning approach for deep face recognition," in *Proceedings of European Conference on Computer Vision (ECCV)*, 2016, pp. 499–515.
- [237] D. P. Kingma and M. Welling, "Auto-encoding variational Bayes," in *Proceedings of International Conference on Learning Representations (ICLR)*, 2014.
- [238] D. P. Kingma, S. Mohamed, D. J. Rezende, and M. Welling, "Semi-supervised learning with deep generative models," in *Advances in Neural Information Processing Systems (NIPS)*, 2014, pp. 3581–3589.
- [239] D. J. Rezende, S. Mohamed, and D. Wierstra, "Stochastic backpropagation and approximate inference in deep generative models," in *Proceedings of International Conference on Machine Learning (ICML)*, 2014.
- [240] C. Doersch, "Tutorial on variational autoencoders," *arXiv preprint arXiv:1606.05908*, 2016.
- [241] E. Wilson, "Backpropagation learning for systems with discrete-valued functions," in *Proceedings of the World Congress on Neural Networks*, vol. 3, 1994, pp. 332–339.
- [242] X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks," in *Proceedings of International Conference on Artificial Intelligence and Statistics (AISTATS)*, vol. 9, 2010, pp. 249–256.
- [243] D. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proceedings of International Conference on Learning Representations (ICLR)*, San Diego, CA, 2015.
- [244] W. Rao and M. W. Mak, "Alleviating the small sample-size problem in i-vector based speaker verification," in *Proceedings of International Symposium on Chinese Spoken Language Processing (ISCSLP)*, 2012, pp. 335–339.
- [245] G. E. Hinton and S. T. Roweis, "Stochastic neighbor embedding," in *Advances in Neural Information Processing Systems (NIPS)*, S. Becker, S. Thrun, and K. Obermayer, Eds., Baltimore, MD: MIT Press, 2003, pp. 857–864.

- [246] H.-H. Tseng, I. E. Naqa, and J.-T. Chien, "Power-law stochastic neighbor embedding," in *Proceedings of International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 2017, pp. 2347–2351.
- [247] J.-T. Chien and C.-H. Chen, "Deep discriminative manifold learning," in *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 2016, pp. 2672–2676.
- [248] K. Chen and A. Salman, "Learning speaker-specific characteristics with a deep neural architecture," *IEEE Transactions on Neural Networks*, vol. 22, no. 11, pp. 1744–1756, 2011.
- [249] J.-T. Chien and C.-W. Hsu, "Variational manifold learning for speaker recognition," in *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 2017, pp. 4935–4939.
- [250] A. Odena, C. Olah, and J. Shlens, "Conditional image synthesis with auxiliary classifier GANs," *arXiv preprint arXiv:1610.09585*, 2016.
- [251] M. Mirza and S. Osindero, "Conditional generative adversarial nets," *arXiv preprint arXiv:1411.1784*, 2014.
- [252] J. Cook, I. Sutskever, A. Mnih, and G. E. Hinton, "Visualizing similarity data with a mixture of maps," in *Proceedings of International Conference on Artificial Intelligence and Statistics (AISTATS)*, 2007, pp. 67–74.
- [253] T. Che, Y. Li, A. P. Jacob, Y. Bengio, and W. Li, "Mode regularized generative adversarial networks," *arXiv preprint arXiv:1612.02136*, 2016.
- [254] M. R. Min, L. Maaten, Z. Yuan, A. J. Bonner, and Z. Zhang, "Deep supervised t -distributed embedding," in *Proceedings of International Conference on Machine Learning (ICML)*, 2010, pp. 791–798.
- [255] D. Palaz, R. Collobert, and M. M. Doss, "Estimating phoneme class conditional probabilities from raw speech signal using convolutional neural networks," in *Proceedings of Annual Conference of International Speech Communication Association (INTERSPEECH)*, 2013, pp. 1766–1770.
- [256] N. Jaitly and G. Hinton, "Learning a better representation of speech soundwaves using restricted Boltzmann machines," in *Proceedings of International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 2011, pp. 5884–5887.
- [257] Z. Tüske, P. Golik, R. Schlüter, and H. Ney, "Acoustic modeling with deep neural networks using raw time signal for LVCSR," in *Proceedings of Annual Conference of International Speech Communication Association (INTERSPEECH)*, 2014.
- [258] D. Palaz, M. Magimai-Doss, and R. Collobert, "End-to-end acoustic modeling using convolutional neural networks for HMM-based automatic speech recognition," *Speech Communication*, 2019.
- [259] Y. Hoshen, R. J. Weiss, and K. W. Wilson, "Speech acoustic modeling from raw multi-channel waveforms," in *Proceedings of International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 2015, pp. 4624–4628.
- [260] D. Palaz, M. Magimai-Doss, and R. Collobert, "Analysis of CNN-based speech recognition system using raw speech as input," in *Proceedings of Annual Conference of International Speech Communication Association (INTERSPEECH)*, 2015, pp. 11–15.
- [261] T. N. Sainath, R. J. Weiss, A. Senior, K. W. Wilson, and O. Vinyals, "Learning the speech front-end with raw waveform CLDNNs," in *Proceedings of Annual Conference of International Speech Communication Association (INTERSPEECH)*, 2015.

- [262] T. N. Sainath, O. Vinyals, A. Senior, and H. Sak, "Convolutional, long short-term memory, fully connected deep neural networks," in *Proceedings of International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 2015, pp. 4580–4584.
- [263] C. Zhang, K. Koishida, and J. H. Hansen, "Text-independent speaker verification based on triplet convolutional neural network embeddings," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 26, no. 9, pp. 1633–1644, 2018.
- [264] C. Zhang and K. Koishida, "End-to-end text-independent speaker verification with triplet loss on short utterances," in *Proceedings of Annual Conference of International Speech Communication Association (INTERSPEECH)*, 2017, pp. 1487–1491.
- [265] J. S. Chung, A. Nagrani, and A. Zisserman, "Voxceleb2: Deep speaker recognition," in *Proceedings Interspeech*, 2018, pp. 1086–1090.
- [266] G. Bhattacharya, J. Alam, and P. Kenny, "Adapting end-to-end neural speaker verification to new languages and recording conditions with adversarial training," in *Proceedings IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*. IEEE, 2019, pp. 6041–6045.
- [267] Y.-Q. Yu, L. Fan, and W.-J. Li, "Ensemble additive margin softmax for speaker verification," in *Proceedings IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*. IEEE, 2019, pp. 6046–6050.
- [268] S. Wang, Y. Yang, T. Wang, Y. Qian, and K. Yu, "Knowledge distillation for small foot-print deep speaker embedding," in *Proceedings IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*. IEEE, 2019, pp. 6021–6025.
- [269] K. He, X. Zhang, S. Ren, and J. Sun, "Spatial pyramid pooling in deep convolutional networks for visual recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, no. 9, pp. 1904–1916, 2015.
- [270] A. Kurakin, I. Goodfellow, and S. Bengio, "Adversarial machine learning at scale," *arXiv preprint arXiv:1611.01236*, 2016.
- [271] A. Makhzani, J. Shlens, N. Jaitly, and I. J. Goodfellow, "Adversarial autoencoders," *CoRR*, vol. abs/1511.05644, 2015. [Online]. Available: <http://arxiv.org/abs/1511.05644>
- [272] Y. Ganin, E. Ustinova, H. Ajakan, P. Germain, H. Larochelle, F. Laviolette, M. Marchand, and V. Lempitsky, "Domain-adversarial training of neural networks," *Journal of Machine Learning Research*, vol. 17, no. 59, pp. 1–35, 2016.
- [273] J. C. Tsai and J. T. Chien, "Adversarial domain separation and adaptation," in *Proceedings IEEE MLSP*, Tokyo, 2017.
- [274] G. Bhattacharya, J. Monteiro, J. Alam, and P. Kenny, "Generative adversarial speaker embedding networks for domain robust end-to-end speaker verification," in *Proceedings IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*. IEEE, 2019, pp. 6226–6230.
- [275] J. Rohdin, T. Stafylakis, A. Silnova, H. Zeinali, L. Burget, and O. Plchot, "Speaker verification using end-to-end adversarial language adaptation," in *Proceedings IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*. IEEE, 2019, pp. 6006–6010.
- [276] X. Fang, L. Zou, J. Li, L. Sun, and Z.-H. Ling, "Channel adversarial training for cross-channel text-independent speaker recognition," in *Proceedings IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 2019, pp. 6221–6225.
- [277] J. Zhou, T. Jiang, L. Li, Q. Hong, Z. Wang, and B. Xia, "Training multi-task adversarial network for extracting noise-robust speaker embedding," in *Proceedings IEEE Interna-*

- tional Conference on Acoustics, Speech, and Signal Processing (ICASSP)*. IEEE, 2019, pp. 6196–6200.
- [278] Z. Meng, Y. Zhao, J. Li, and Y. Gong, “Adversarial speaker verification,” in *Proceedings IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*. IEEE, 2019, pp. 6216–6220.
- [279] P. S. Nidadavolu, J. Villalba, and N. Dehak, “Cycle-gans for domain adaptation of acoustic features for speaker recognition,” in *Proceedings IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*. IEEE, 2019, pp. 6206–6210.
- [280] L. Li, Z. Tang, Y. Shi, and D. Wang, “Gaussian-constrained training for speaker verification,” in *Proceedings IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*. IEEE, 2019, pp. 6036–6040.
- [281] Y. Tu, M.-W. Mak and J.-T. Chien, “Variational domain adversarial learning for speaker verification,” in *Proceedings of Annual Conference of International Speech Communication Association (INTERSPEECH)*, 2019, pp. 4315–4319.
- [282] S. S. Shapiro and M. B. Wilk, “An analysis of variance test for normality (complete samples),” *Biometrika*, vol. 52, no. 3/4, pp. 591–611, 1965.