

FOCAL KL-DIVERGENCE BASED DILATED CONVOLUTIONAL NEURAL NETWORKS FOR CO-CHANNEL SPEAKER IDENTIFICATION

Shuai Wang, Yanmin Qian[†], Kai Yu

Key Laboratory of Shanghai Education Commission for Intelligent Interaction and Cognitive Engineering
SpeechLab, Department of Computer Science and Engineering
Brain Science and Technology Research Center
Shanghai Jiao Tong University, Shanghai, China
{wsstriving@gmail.com, yanminqian@tencent.com, kai.yu@sjtu.edu.cn}

ABSTRACT

Recognizing the identities of multiple talkers via their overlapped speech is a challenging task, it is also one main difficulty for the “cocktail party problem”. In this paper, a novel dilated convolutional neural network with a focal KL-divergence loss function is proposed to tackle this problem. During training, relative loss for the well-classified samples is automatically reduced and consequently more attention is paid to the hard samples. The use of the focal KL-divergence loss function leads to more stable training and improved testing performance. Furthermore, a post processing of assigning different frames with different weights is also adopted and leads to further improvement. The proposed framework can be easily extended from 2-talker to 3-talker speaker identification scenario. Experiments on the artificially generated RSR2015 multi-talker mixed corpus show that the proposed approach can improve multi-talker speaker identification significantly.

Index Terms— co-channel speaker identification, convolutional neural networks, focal loss, dilated convolution.

1. INTRODUCTION

Co-channel speaker identification (co-channel SID) aims to recognize the identities of multiple talkers when they speak simultaneously, it is one main challenge of the “Cocktail Party Problem” [1, 2]. Although the state-of-the-art speaker identification (SID) systems can achieve impressive accuracy in the single-talker scenario, processing highly overlapped speech is still a very challenging task in the speaker recognition research area.

Before deep learning era, Gaussian Mixture Model (GMM) based approaches were usually adopted. In [3], authors proposed to treat co-channel SID and speech separation as an iterative process, adapted GMM and KL-Divergence (KLD) scores were fused to further improve the performance. In [4], the most probable speaker is firstly selected and paired with other possible speakers and the combined models are used for co-channel SID.

[†]Yanmin Qian and Kai Yu are the corresponding authors and now Yanmin Qian is with Tencent AI Lab, Tencent, Bellevue, WA, USA.

This work has been supported by the National Key Research and Development Program of China under Grant No.2017YFB1002102 and the China NSFC projects (No. U1736202 and No. 61603252). Experiments have been carried out on the PI supercomputer at Shanghai Jiao Tong University.

Motivated by the success of Deep learning for related tasks, e.g. speech recognition [5, 6, 7], Deep Neural Networks (DNN) were also tried for speaker recognition. DNN-based speaker recognition systems, such as d -vector [8], j -vector [9, 10] and end-to-end frameworks [11, 12, 13, 14] were proposed in recent years. However, there are few studies utilizing deep learning in the co-channel SID task. In [15], authors treated the co-channel SID as a multi-class classification problem and utilize DNN as the classifier. A basic DNN model can achieve nearly 100.0% accuracy on the simple SSC dataset, a similar improvement can also be observed on the artificially generated SRE overlapped speech [15], which is harder.

Convolutional Neural Network(CNN) is well-known for its incredible capability of learning structured features and it is widely used in many fields such as image recognition [16] and speech recognition [17]. In this paper, CNN is utilized in our system, showing the superiority over DNN on the co-channel SID problem. To enlarge the reception field, dilated convolution is adopted. Moreover, we propose an enhanced version of KL-Divergence, which is named Focal-KLD for model optimization: the loss for simple samples is reduced and that for those hard samples is increased. This new loss function stabilizes the training process and achieves a better performance. Finally, post filtering is applied to get an additional gain. Experiments were carried out on the relatively simple speech separation challenge (SSC) corpus as well as the harder artificially generated RSR multi-talker mixed corpus. Significant improvements are observed by the proposed approaches.

The rest of this paper is organized as follows, Section 2 briefly reviews the DNN based co-channel SID system, and Section 3 describes the proposed focal kl-divergence based dilated convolution neural network system. Section 4 introduces the experimental setups and results comparison, and Section 5 concludes this paper.

2. DNN BASED CO-CHANNEL SID

Since co-channel SID is a really challenging task, existing methods only consider the close-set scenario, in which all speakers in the inference phase are known to the training set. Under this condition, in [15, 18], co-channel SID is formulated as a multi-class (class number is known) classification problem and DNN is employed as the classifier. The flow chart is shown in Figure 1. Given artificially generated multi-talker mixed frame-level features as the input, soft training labels, representing the probabilities of underlying speak-

ers to generate the current frame, are used as the targets. The sum of the probabilities of target speakers equals one, whereas the other speakers have zero probabilities. Soft labels are computed using the frame-level energy ratio, more details can be found in [15, 18]. In the evaluation phase, frame-level scores are aggregated at the utterance level. Given a test utterance \mathbf{O} consisting of T frames $\mathbf{o}_1, \mathbf{o}_1 \cdots \mathbf{o}_T$, the utterance-level probability for speaker s will be computed as,

$$\mathcal{J}(s) = \frac{1}{T} \sum_{t=1}^T P(s|\mathbf{o}_t) \quad (1)$$

where $P(s|\mathbf{o}_t)$ represents the probability of frame \mathbf{o}_t comes from speaker s . The predicted speaker identities are obtained by selecting the top k speakers with the largest probabilities.

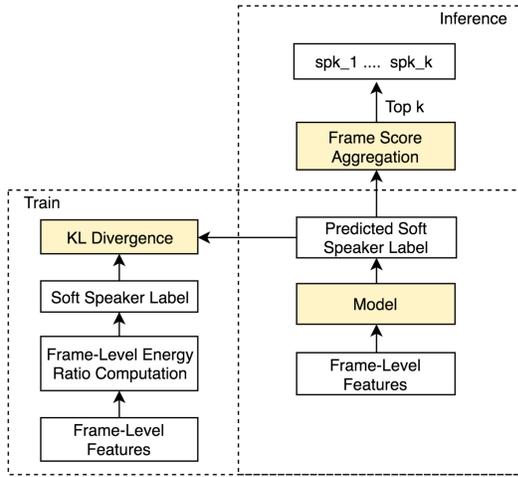


Fig. 1. Co-channel SID system using deep models

3. FOCAL KL-DIVERGENCE BASED DILATED CONVOLUTIONAL NEURAL NETWORK

Following the framework depicted in Figure 1, we improve the system from three aspects. First, instead of DNN, CNN is adopted as the learning machine, it's more capable of learning structured features. Moreover, we adopt dilated convolution in the first several layers in our proposed system for multi-talker SID, allowing a better leverage of the context information. Second, an enhanced version of KL-Divergence (Focal-KLD) is proposed to substitute the normal KLD loss. Finally, a *post filtering* operation is performed instead of simple averaging in the frame score aggregation phase.

3.1. Dilated Convolution

A dilated convolution is a convolution where the filter is applied over an area larger than its length by skipping input values with a step [19, 20], thus allowing the network to operate on a coarser scale than a normal convolution and supports an exponential expansion on the receptive field. A dilated convolution can be regarded as a normal convolution with a larger filter size, where some entries are neglected in the computation. Figure 2 gives an example of 2-dilated convolution with a 3×3 filter.

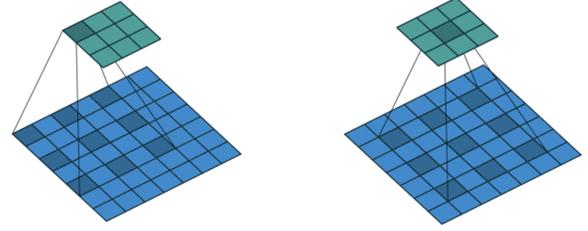


Fig. 2. Convolution using a 3×3 filter with dilation of 2.

3.2. Focal Kullback-Leibler Divergence Loss

3.2.1. Kullback-Leibler Divergence Loss

Different from the objective function used in d -vector paradigm in single-talker recognition task, as a multi-class classification task, Kullback-Leibler Divergence (KLD), which measures the distance between two probability distributions, is usually used as the loss function in neural network optimization. For instance, KL-Divergence is added as a regularization to the adaption criterion in [21]. Recently, transfer learning with teacher-student training has received more and more attention [22, 23, 24, 25], which also uses the KL-Divergence as objective function. The frame-wise KL-Divergence is defined as

$$\text{KLD}(\theta; \mathbf{o}, \mathbf{y}) = \sum_{i=1}^D p_{\text{ref}}(y_i|\mathbf{o}) \log \frac{p_{\text{ref}}(y_i|\mathbf{o})}{p_{\theta}(y_i|\mathbf{o})} \quad (2)$$

where $p_{\theta}(y_i|\mathbf{o})$ is the posterior for observation \mathbf{o} computed by the trained model parametrized with θ , $p_{\text{ref}}(y_i|\mathbf{o})$ is the reference target soft label. D is the dimension of \mathbf{y} , representing the number of speakers. It should be noted that cross-entropy is adopted as the loss function in the previous work [18], which is actually equivalent to KLD in this setting. Equation 2 can be re-written as

$$\text{KLD}(\theta; \mathbf{o}, \mathbf{y}) = \text{CE}(\theta; \mathbf{o}, \mathbf{y}) + \text{const} \quad (3)$$

$$\text{CE}(\theta; \mathbf{o}, \mathbf{y}) = - \sum_{i=1}^D p_{\text{ref}}(y_i|\mathbf{o}) \log p_{\theta}(y_i|\mathbf{o}) \quad (4)$$

$$\text{const} = \sum_{i=1}^D p_{\text{ref}}(y_i|\mathbf{o}) \log p_{\text{ref}}(y_i|\mathbf{o}) \quad (5)$$

Since the second part of Equation 3 is constant given the training set, it's equivalent to optimize either loss function. Because cross entropy is used in the single-class classification, KLD is adopted here to eliminate ambiguity.

3.2.2. Focal Kullback-Leibler Divergence Loss

Focal loss was first proposed in [26], it adds a weighting factor $(1 - P(y_{\text{target}}|\mathbf{o})^{\gamma})$ to the standard KLD criterion. Setting $\gamma > 0$ reduces the relative loss for well-classified samples (simple samples), and increases more focus on the harder and misclassified ones. In this work, we designed the focal version of normal KLD and applied it in for multi-talker SID. It is further referred to as Focal KLD (FKLD).

$$\text{FKLD}(\theta; \mathbf{o}, \mathbf{y}) = w \cdot \text{KLD}(\theta; \mathbf{o}, \mathbf{y}) \quad (6)$$

$$w = (1 + \alpha - \sum_{i|P_{\text{ref}}(y_i|\mathbf{o}) \neq 0} P_{\theta}(y_i|\mathbf{o})^{\gamma}) \quad (7)$$

where α and γ are two hyper-parameters, controlling the decaying extent of the loss. These two parameters can be fixed through the training process, or be changed according to the specific training condition (e.g. changing by the training epochs), and details will be discussed in Section 4.4.

3.3. Post Filtering

The simplest way to perform frame score aggregation during the inference is to average the frame-level probabilities, shown in Equation 1. Although this method is simple and effective, it treats all frames equally. We propose to assign different frames with individual weights to boost the performance, which is named post filtering. Equation 1 is changed to

$$\mathcal{J}(s) = \frac{1}{T} \sum_{t=1}^T (\mathbf{w}_t)^\beta \cdot p(s|\mathbf{o}_t) \quad (8)$$

where $\mathbf{w}_t = \max_{s \in S} p(s|\mathbf{o}_t)$, representing the largest probability of this frame, and β is an adjustable hyper-parameter. As shown in Equation 8, for each frame-level probability vector, the weight it earns is larger when it's more likely generated by a single speaker. We can think that it gives more confidence on the non-overlapped speech segments and less confidence on the overlapped ones.

4. EXPERIMENTS

4.1. System Description

4.1.1. Data Preparing and Evaluation Metric

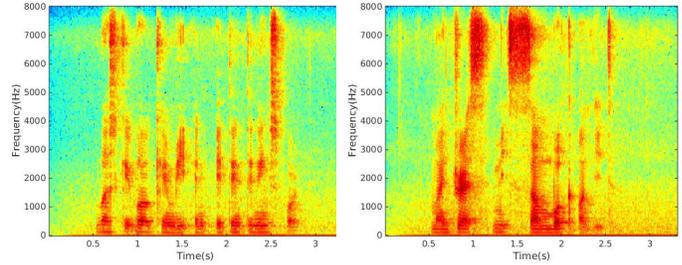
In this paper, in addition to the standard SSC corpus[27], an artificially generated multi-talker SID corpus based on RSR2015 is utilized, including two- and three-talker mixed speech scenarios (called RSR-2mix and RSR-3mix, respectively). Taking the two-talker speech as an example, given the clean utterances in the original corpus, the utterance-pairs are randomly chosen and the shorter utterance is padded to match the length of the longer one. The selected utterances are then mixed with the equal energy ratio, i.e. 0dB SNR. Figure 3 illustrates an exemplary spectrogram of the original clean single talker speech and the mixed two-talker speech. Due to the large overlapped area in the mixed speech, the co-channel mixed speech is much challenging for SID than the single-talker speech. The three-talker mixed speech data is generated following similar procedure, with the equal energy ratio within three speakers.

40-dimensional Fbank is used as features in all experiments, large chunks of silence are removed using an energy-based VAD, after which Cepstral Mean Subtraction (CMS) is performed.

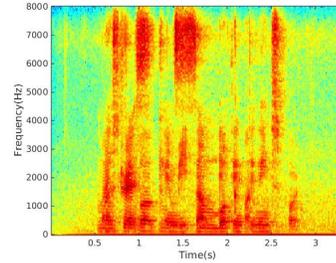
Prediction accuracy is used as the evaluation metric. A test case is only correct **when all potential speakers are correctly predicted**.

4.1.2. Baseline DNN System

As the previous work in [18], DNN with normal KLD is employed as the baseline system. It contains 4 hidden layers, each consists of 512 nodes. ReLU is used as the activation function, and the model initialization follows the settings in [28]. SGD is used as the optimization method with the momentum set as 0.9.



(a) First speaker clean speech (b) Second speaker clean speech



(c) Two-talker mixed speech under 0dB SNR condition

Fig. 3. Spectrogram comparison of single-talker clean speech and two-talker mixed speech in RSR-2mix dataset

4.1.3. Proposed Dilated CNN system

To make the model scale comparable, the CNN used in this work contains 3 convolution layers with 1 fully-connected layer. No pooling layer is involved. Different paddings (on both sides) are chosen to make the feature map size unchanged. Table 1 gives the detailed configuration, the size of the output layer corresponds to number of speakers of the dataset(34 for SSC, 50 for RSR dataset).

Table 1. Dilated CNN Configuration

Input	40 × 11 Feature Map (11-frame context)			
Conv Layers	filter size	padding	dilation	channels
1	5×5	2	1	2
2	3×3	1	1	4
3	3×3	2	2	6
FC Layer	512 nodes			
Output	34 nodes (SSC)		50 nodes (RSR)	

4.2. Validation Experiments on SSC Corpus

The Speech Separation Challenge (SSC) Corpus[27] contains 17000 training utterances from 34 speakers. Each training utterance is generated following a fixed pattern: *command, color, preposition, letter, number* and *adverb*. Test utterances are mixed single-channel speech generated from the same 34 speakers, but using different utterances from the training set. Though the test set comprises six TIRs from -9 dB to 6 dB, we only performed on the 0dB case in our experiments, which contains 600 test cases. The results in Table 2 show that our baseline system (using DNN with KLD) achieves 100.0% accuracy

in the 2-talker case, which is comparable to (even better than) the previous published result on this task. Considering that the accuracy on SSC has been nearly perfect, we will change to another harder corpus to better evaluate the proposed approaches.

Table 2. Accuracy (%) comparison on SSC Dataset

Systems	2 talkers
DNN+KLD in [15]	99.8
DNN+KLD in our work	100

4.3. Experiments on multi-talker RSR 2015 Corpus

4.3.1. Multi-talker dataset design

Since the SSC task is too simple to get useful conclusions, we carried out more experiments on the artificially generated multi-talker RSR corpus. The original RSR2015 corpus is not designed for co-channel speaker identification task, so we artificially generated a multi-talker corpus based on the RSR 2015 part 1. 50 speakers (25 males and 25 females) are randomly selected. In the 2-talker experiments, there can be 1225 (50*49/2) speaker pairs. For each speaker pair, we randomly select one utterance from each speaker and generate one co-channel utterance. Total 20 co-channel utterances are generated for each speaker pair, resulting in 24500 training utterances. We follow the same procedure to generate the evaluation utterances, 4 co-channel utterances for each speaker pair, resulting in 4900 test cases. For the 3-talker scenario, there are 19600 (50*49*48/6) speaker triplets. For each speaker triplet, we generate 3 co-channel utterances, leading to 58800 training utterances. Following the same procedure, 10000 testing utterances are randomly generated. It should be noted that all the utterances used for generating test cases are not included in the training set.

4.3.2. Dilated CNN with focal KL-Divergence

The convergence curves of the baseline KLD-DNN and the proposed FKLD-based Dilated CNN are shown in Figure 4. It is observed that the proposed FKLD-based dilated CNN converges faster and better than the baseline. The accuracy on the validation set achieves 68.5% after the first training epoch for the dilated CNN system, while only 37.3% is obtained for the DNN.

Table 3 shows the recognition results of the baseline DNN system and the systems using the proposed approaches. Focal refers to deep models trained with the enhanced Focal-KLD loss, and PF means using the post filtering processing in the evaluation phase (described in Section 3.3). As shown as Table 3, the proposed dilated CNN structure outperforms the baseline DNN in both 2-talker and 3-talker scenarios. By replacing the normal KL-Divergence loss to the proposed focal KL-Divergence can improve the accuracy consistently on both conditions. The accuracy for 2- and 3-talker conditions are increased from 87.16% and 47.79% to 91.31% and 55.74% respectively by the dilated CNN structure with focal KLD.

As described in Section 3.2, a weighting function termed as *post filtering* can be adopted in the frame-level score aggregation phase, resulting in another significant improvement. The accuracies are further increased to 92.47% and 55.83% for the 2- and 3-talker conditions respectively.

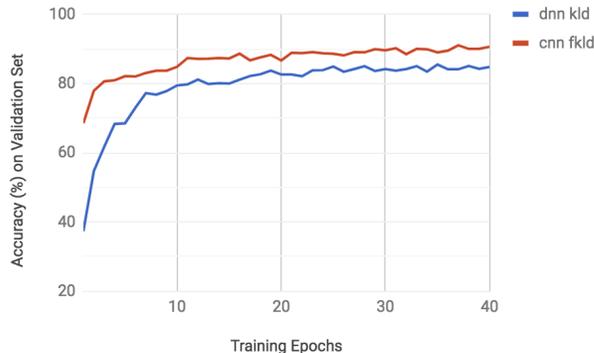


Fig. 4. Convergence curve comparison of the baseline KLD-DNN and the proposed FKLD-Dilated CNN on the generated RSR-2mix dataset

Table 3. Accuracy (%) comparison on multi-talker RSR Dataset

Models	Focal	PF	2 talkers	3 talkers
DNN	×	×	87.16	47.79
	✓	×	88.59	51.91
	✓	✓	89.24	52.51
CNN	×	×	88.65	50.86
	✓	×	91.31	55.74
	✓	✓	92.47	55.83

4.4. The hyper-parameter setting

There are several hyper parameters in the proposed approach, including α , γ in Equation 4 and β in Equation 5. α is suggested to be set between 0.0 and 0.5, forbidding the decay factor to be either too small or too large, and γ can be a fixed value or a changed one. Different from the γ usage with fixed to 2.0 in the previous work for image recognition [26] actually we increase γ gradually in this work, i.e. paying more and more attentions on the hard samples during the training process. To be more specific, in the 2-talker experiments, α is set to 0.3 and γ is set to ($\#epoch/10$), and in the 3-talker experiments, α is set to 0.5 and γ is set to ($\#epoch/10$). β in Equation 8 is set to 2.0 and 1.0 for the 2- and 3-talker conditions respectively.

5. CONCLUSION

We proposed a dilated CNN-based framework to tackle the challenging Co-channel multi-talker SID problem. Experiments were carried out on both standard SSC and artificially generated multi-talker RSR corpus. Compared to the previous work using DNN, the proposed dilated CNN works well on encoding the speech features in the co-channel condition and shows better performance. Moreover, we proposed a new Focal Kullback-Leibler Divergence (FKLD) loss function for the model optimization, which reduces loss on the well-classified (simple) samples and pays more attention on the mis-classified (hard) samples. This new loss function can get a significant improvement compared to the normal KLD. Finally a post filtering operation is performed during the inference phase to further refine the system performance.

6. REFERENCES

- [1] Dong Yu, Xuankai Chang, and Yanmin Qian, “Recognizing multi-talker speech with permutation invariant training,” in *INTERSPEECH*, 2017, pp. 2456–2460.
- [2] Yanmin Qian, Xuankai Chang, and Dong Yu, “Single-channel multi-talker speech recognition with permutation invariant training,” *arXiv preprint arXiv:1707.06527*, submitted to *TASLP*, 2017.
- [3] Pejman Mowlaei, Rahim Saeidi, Z-H Tan, Mads Græsbøll Christensen, Pasi Fränti, and Søren Holdt Jensen, “Joint single-channel speech separation and speaker identification,” in *IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP)*. IEEE, 2010, pp. 4430–4433.
- [4] Peng Li, Yong Guan, Shijin Wang, Bo Xu, and Wenju Liu, “Monaural speech separation based on MAXVQ and CASA for robust speech recognition,” *Computer Speech & Language*, 2010.
- [5] Geoffrey Hinton, Li Deng, Dong Yu, George E Dahl, Abdelrahman Mohamed, Navdeep Jaitly, Andrew Senior, Vincent Vanhoucke, Patrick Nguyen, Tara N Sainath, et al., “Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups,” *IEEE Signal Processing Magazine*, vol. 29, no. 6, pp. 82–97, 2012.
- [6] Frank Seide, Gang Li, and Dong Yu, “Conversational speech transcription using context-dependent deep neural networks,” in *Twelfth Annual Conference of the International Speech Communication Association*, 2011.
- [7] George E Dahl, Dong Yu, Li Deng, and Alex 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.
- [8] Ehsan Variani, Xin Lei, Erik McDermott, Ignacio Lopez Moreno, and Javier Gonzalez-Dominguez, “Deep neural networks for small footprint text-dependent speaker verification,” in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2014, pp. 4052–4056.
- [9] Nanxin Chen, Yanmin Qian, and Kai Yu, “Multi-task learning for text-dependent speaker verification,” in *INTERSPEECH*, 2015.
- [10] Yuan Liu, Yanmin Qian, Nanxin Chen, Tianfan Fu, Ya Zhang, and Kai Yu, “Deep feature for text-dependent speaker verification,” *Speech Communication*, vol. 73, pp. 1–13, 2015.
- [11] Georg Heigold, Ignacio Moreno, Samy Bengio, and Noam Shazeer, “End-to-end text-dependent speaker verification,” in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2016, pp. 5115–5119.
- [12] Chao Li, Xiaokong Ma, Bing Jiang, Xiangang Li, Xuewei Zhang, Xiao Liu, Ying Cao, Ajay Kannan, and Zhenyao Zhu, “Deep speaker: an end-to-end neural speaker embedding system,” *arXiv preprint arXiv:1705.02304*, 2017.
- [13] David Snyder, Pegah Ghahremani, Daniel Povey, Daniel Garcia-Romero, Yishay Carmiel, and Sanjeev Khudanpur, “Deep neural network-based speaker embeddings for end-to-end speaker verification,” in *Spoken Language Technology Workshop (SLT), 2016 IEEE*. IEEE, 2016, pp. 165–170.
- [14] Chunlei Zhang and Kazuhito Koishida, “End-to-end text-independent speaker verification with triplet loss on short utterances,” *INTERSPEECH*, pp. 1487–1491, 2017.
- [15] Xiaojia Zhao, Yuxuan Wang, and DeLiang Wang, “Cochannel speaker identification in anechoic and reverberant conditions,” *IEEE/ACM Transactions on Audio, Speech and Language Processing (TASLP)*, vol. 23, no. 11, pp. 1727–1736, 2015.
- [16] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton, “Imagenet classification with deep convolutional neural networks,” in *Advances in neural information processing systems*, 2012, pp. 1097–1105.
- [17] Yanmin Qian, Mengxiao Bi, Tian Tan, and Kai Yu, “Very deep convolutional neural networks for noise robust speech recognition,” *IEEE/ACM Transactions on Audio, Speech and Language Processing*, vol. 24, no. 12, pp. 2263–2276, 2016.
- [18] Xiaojia Zhao, Yuxuan Wang, and DeLiang Wang, “Deep neural networks for cochannel speaker identification,” in *IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP)*. IEEE, 2015, pp. 4824–4828.
- [19] Aaron van den Oord, Sander Dieleman, Heiga Zen, Karen Simonyan, Oriol Vinyals, Alex Graves, Nal Kalchbrenner, Andrew Senior, and Koray Kavukcuoglu, “Wavenet: A generative model for raw audio,” *arXiv preprint arXiv:1609.03499*, 2016.
- [20] Fisher Yu and Vladlen Koltun, “Multi-scale context aggregation by dilated convolutions,” *arXiv preprint arXiv:1511.07122*, 2015.
- [21] Dong Yu, Kaisheng Yao, Hang Su, Gang Li, and Frank Seide, “KL-divergence regularized deep neural network adaptation for improved large vocabulary speech recognition,” in *IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP)*. IEEE, 2013, pp. 7893–7897.
- [22] Yevgen Chebotar and Austin Waters, “Distilling knowledge from ensembles of neural networks for speech recognition,” in *INTERSPEECH*, 2016, pp. 3439–3443.
- [23] Jeremy HM Wong and Mark JF Gales, “Sequence student-teacher training of deep neural networks,” 2016.
- [24] Geoffrey Hinton, Oriol Vinyals, and Jeff Dean, “Distilling the knowledge in a neural network,” *arXiv preprint arXiv:1503.02531*, 2015.
- [25] Jinyu Li, Rui Zhao, Jui-Ting Huang, and Yifan Gong, “Learning small-size dnn with output-distribution-based criteria,” in *INTERSPEECH*, 2014.
- [26] Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, and Piotr Dollár, “Focal loss for dense object detection,” *arXiv preprint arXiv:1708.02002*, 2017.
- [27] T. Lee M. P. Cooke, “The 2006 speech separation challenge,” 2006.
- [28] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun, “Delving deep into rectifiers: Surpassing human-level performance on imagenet classification,” *arXiv preprint arXiv:1502.01852*, 2015.