

张杨
特殊学术专长推免生
佐证材料

姓名：张杨

学号：23080309

专业：通信工程

学院：通信工程学院

一、四六级

1. 四级:

全国大学英语四级考试

成绩报告单





姓名: 张杨

学校: 杭州电子科技大学

院系: 通信工程学院

身份证号: 331081200412123085

准考证号: 330371232113830

考试时间: 2023年12月

成绩报告单编号: 232133037003226

笔试

总分	听力 (35%)	阅读 (35%)	写作和翻译 (30%)
547	174	205	168

口试

准考证号: --

考试时间: --

成绩: --



校验码: E54I OPXI 9MB2 DVXM



说明

1. 全国大学英语四、六级考试（CET）是由教育部主办的全国统一考试，考试对象为在校大学生。考试内容
包括听、说、读、写、译等语言技能。
2. CET笔试考试时间为每年6月和12月；CET口试考试
时间为每年5月和11月。
3. 考生可登录中国教育考试网（www.neea.edu.cn）查
询、下载电子成绩报告单或自行办理纸质成绩证明。
电子成绩报告单和纸质成绩证明与纸质成绩报告单具
有同等效力。

大学英语四级口语考试能力描述

优秀	能用英语就熟悉的话题进行有效的交流； 能清晰地叙述或描述一般性事件和现象。 语言表达清楚连贯。
良好	能用英语就熟悉的话题进行交流；能叙述 或描述一般性事件和现象。语言表达基本 准确。
合格	能用英语就熟悉的话题进行简单交流；能 简单叙述或描述一般性事件和现象。

2. 六级:

全国大学英语六级考试

成绩报告单





姓名: 张杨

学校: 杭州电子科技大学

院系: 通信工程学院

身份证号: 331081200412123085

准考证号: 330371261222712

考试时间: 2026年6月

成绩报告单编号: 261233037004226

笔试

总分	听力 (35%)	阅读 (35%)	写作和翻译 (30%)
494	156	164	174

口试

准考证号: --

考试时间: --

成绩: --



校验码: 3W98 K2IQ ME1O OSV6



说明

1. 全国大学英语四、六级考试（CET）是由教育部主办的全国统一考试，考试对象为在校大学生。考试内容
包括听、说、读、写、译等语言技能。
2. CET笔试考试时间为每年6月和12月；CET口试考试
时间为每年5月和11月。
3. 考生可登录中国教育考试网（www.neea.edu.cn）查
询、下载电子成绩报告单或自行办理纸质成绩证明。
电子成绩报告单和纸质成绩证明与纸质成绩报告单具
有同等效力。

大学英语六级口语考试能力描述

优秀	能用英语就一般性话题清晰地阐述自己的 观点，明确地表达自己的态度；能开展深 入的讨论，发表具有一定深度的见解。语 言表达适切，自然流畅。
良好	能用英语就一般性话题阐述自己的观点， 表明自己的态度；能开展较深入的讨论。 语言表达准确连贯。
合格	能用英语就一般性话题进行交流；能参与 讨论。语言表达基本准确。

二、科研积分—学术论文—一级及以上期刊（ccf-b 会议）（8 分）

指导老师签字：江文斌，情况属实

日期：2026 年 8 月 31 日

1. ccf 分级(“计算机图形学与多媒体”分类 B 类推荐国际学术会议)

二、B 类

序号	会议名称	出版商	网址
1	ICMR	ACM SIGMM International Conference on Multimedia Retrieval	ACM http://dblp.uni-trier.de/db/conf/icmr/
2	I3D	ACM SIGGRAPH Symposium on Interactive 3D Graphics and Games	ACM http://dblp.uni-trier.de/db/conf/i3d/
3	SCA	ACM SIGGRAPH Symposium on Computer Animation	ACM http://dblp.uni-trier.de/db/conf/sca/index.html
4	DCC	Data Compression Conference	IEEE http://dblp.uni-trier.de/db/conf/dcc/
5	Eurographics	Annual Conference of the European Association for Computer Graphics	Wiley/Blackwell http://dblp.uni-trier.de/db/conf/eurographics/
6	EuroVis	Eurographics Conference on Visualization	ACM http://dblp.uni-trier.de/db/conf/eurovis/
7	SGP	Eurographics Symposium on Geometry Processing	Wiley/Blackwell http://dblp.uni-trier.de/db/conf/sgp/
8	EGSR	Eurographics Symposium on Rendering	Wiley/Blackwell http://dblp.uni-trier.de/db/conf/egsr/
9	ICASSP	IEEE International Conference on Acoustics, Speech and Signal Processing	IEEE http://dblp.uni-trier.de/db/conf/icassp/
10	ICME	IEEE International Conference on Multimedia & Expo	IEEE http://dblp.uni-trier.de/db/conf/icme/
11	ISMIR	International Symposium on Mixed and Augmented Reality	IEEE/ACM http://dblp.uni-trier.de/db/conf/ismir/
12	PO	Pacific Conference on Computer Graphics and Applications	Wiley/Blackwell http://dblp.uni-trier.de/db/conf/po/index.html
13	SPM	Symposium on Solid and Physical Modeling	SIAM/Elsevier http://dblp.uni-trier.de/db/conf/spm/
14	INTER-SPEECH	Conference of the International Speech Communication Association	ISCA http://dblp.uni-trier.de/db/conf/interpeech/index.html

2. 论文接收邮件

mail.hdu.edu.cn/static/sirius-web/?version=17878982C

收件箱 ICASSP 2026: ... ICASSP 2026: A...

ICASSP 2026: Review Results [Paper #17075]

papers <papers@2026.ieeeicassp.org> 2026-01-18 03:45

收件人: 我, Wenbin Jiang, Zhen Wang, 吴开颖, 张雯, Fei Wen 收起

附件: papers papers@2026.ieeeicassp.org

发件人: 江文斌 <23090309@hdu.edu.cn>, Wenbin Jiang <wbjiang@hdu.edu.cn>, Zhen Wang <2319841604@qq.com>, 吴开颖 <23080802@hdu.edu.cn>, 张雯 <23080911@hdu.edu.cn>, Fei Wen <wenfei@hdu.edu.cn>

时间: 2026-01-18 03:45:59

● 英语 ● 翻译为 简体中文 ● 翻译附件

Dear Yang Zhang, Wenbin Jiang, Zhen Wang, KaiYing Wu, Wen Zhang, Fei Wen,

Paper ID: 17075

Title: HyFlowSE: Hybrid End-to-End Flow-Matching Speech Enhancement via Generative-Discriminative Learning

Congratulations!

We are pleased to inform you that the above-listed manuscript has been accepted for presentation at IEEE ICASSP 2026 (<https://2026.ieeeicassp.org>), which will be held as an in-person event in Barcelona, Spain, from May 4 to 8, 2026.

Over the past few months, the IEEE Signal Processing Society's Technical Committees worked very hard reviewing the 11,120 papers submitted to ICASSP 2026. This was the largest and most comprehensive ICASSP in IEEE SPS history. Expert reviewers evaluated submissions in their areas of expertise, resulting in a highly competitive selection process with an acceptance rate of about 40%.

Thank you for contributing to IEEE ICASSP 2026! To celebrate your paper acceptance, we've created a shareable graphic you can use to promote your work and invite your network to attend. Use this link to spread the word across LinkedIn, X, Slack, WhatsApp, email, and more: <https://app.gleanin.com/share/c/43733>

The reviewers' comments on your paper are attached at the end of this email. If appropriate, feedback from the leadership group of the Technical Committee (TC) is also included after the reviewers' comments. These comments served as part of the basis for the final decision. Please consider the reviewers' comments and the TC feedback to strengthen your final paper.

ICASSP 2026
Barcelona

51st IEEE International Conference on
Acoustics, Speech and Signal Processing

ICASSP 2026

May 3-8, 2026, Barcelona, Spain

Where Signals Meet Intelligence

URGENT!

General Chairs
Ana I. Pérez-Neira
Xavier Mestre
Technical Program Chairs
Markus Rupp
Christian Jutten
Tülay Adalı
Finance Chair
Jesús Gómez
Special Session Chairs
Arnold Lee Swindlehurst
Isabel Trancoso
Tutorial & Short Courses Chairs
Maria Sabrina Greco
Phillip Regalia
Plenaries Chairs
Pier Luigi Dragotti
Petar Djuric
Workshops Chair
Tülay Adalı
Industrial TPC Chair
Fa-Long Luo
Operations Chair
Adrian Agapin
Challenges Chair
Jordi Villà Valls
Exhibit Chair
Adriano Pastore
Industry Keynote Chair
Mike Polley
Industry Panel Chair
Pu (Perry) Wang
Industry Expert Sessions Chair
Yang Lei
Show and Tell Chair
Ivan Tashev
Entrepreneurship Forum Chair
Elisabharby Ambiairajah
Student Activities Chairs
Martin Haardt
Mónica Bugallo

Sunday, 18 January 2026

First Name: [Yang], Surname: [Zhang]
Department: School of Communication Engineering
University: Hangzhou Dianzi University
Phone: 13282869298
Date of Birth: 12 December 2004
Gender: Female
City: Hangzhou
State/Province: Zhejiang
Postal Code: 310018
Country/Region: China
Passport Issue: China

Dear Yang Zhang,

Regarding presentation of the following **accepted paper**
HyFlowSE: Hybrid End-to-End Flow-Matching Speech Enhancement via Generative-Discriminative Learning

On behalf of IEEE ICASSP 2026, I invite you to join us for the upcoming conference.

We are pleased to inform you that your submission has been accepted for presentation at the 2026 IEEE International Conference on Acoustics, Speech, and Signal Processing (IEEE ICASSP 2026) in Barcelona, Spain, during 3-8 May 2026. [ICASSP is the world's largest and most comprehensive technical conference focused on signal processing and its applications. It offers a variety of technical program presenting all the latest development in research and technology in the industry that attracts thousands of professionals annually.

Note that at least one author must register for the conference at an author rate by 13 February 2026. Failure to have such registration will result in removal of the paper from the Technical Program. Note that your paper must be presented and questions answered to comply with the No Show Policy.

This letter of invitation is not an offer to contribute to the financial costs of registration, travel to and from the conference, or subsistence while at the conference or traveling. Delegates and authors are responsible for the entire travel costs, to and from their home countries, including hotel expenses during the conference period.

You may visit our website for more updated information
<https://2026.ieeeicassp.org/>.

Again, congratulations! We look forward to welcoming you to IEEE ICASSP 2026.

Sincerely,

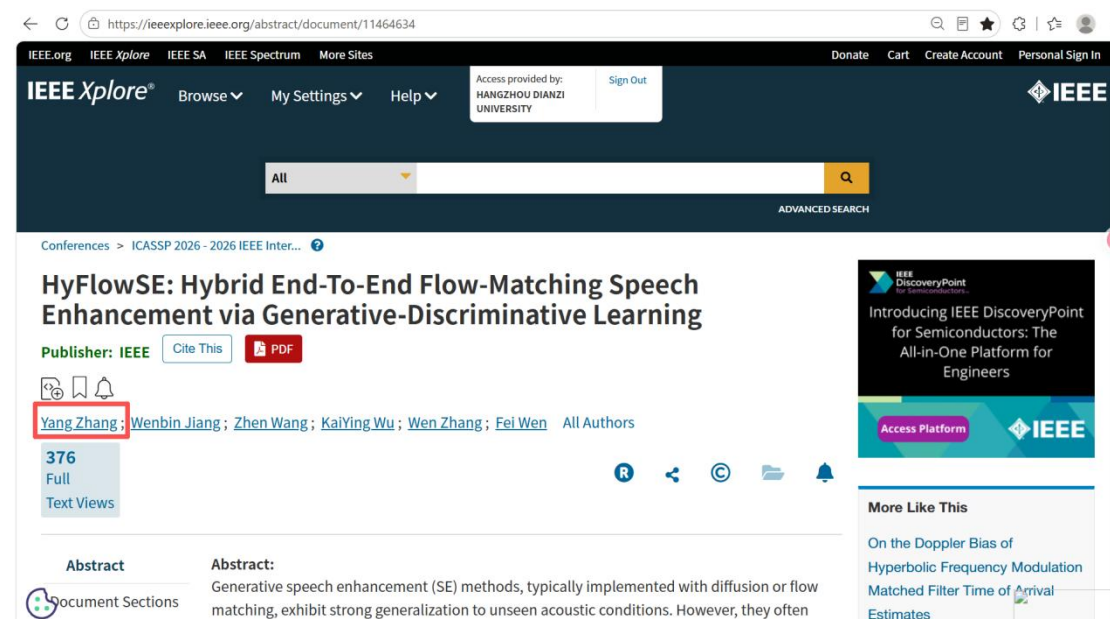
Xavier Mestre
General Chair IEEE
ICASSP 2026
generalchair@2026.
ieeeicassp.org

Yang Zhang
Billete Canton
ICASSP 2026 Conference Manager
Conference Management Office
2711 Pierre Pl
College Station, Texas, 77745
United States
Phone: +1 979-846-6800
Email: registration@2026.ieeeicassp.org

[Event/ICASSP 2026 Paper/17875 Auth/62783 Gen/2026-01-18T02:29:47+00:00]

会议论文集网址: <https://ieeexplore.ieee.org/xpl/conhome/11460365/proceeding>

5. 论文在线发表界面



HyFlowSE: Hybrid end-to-end flow-matching speech enhancement via generative-discriminative learning



Authors Yang Zhang, Wenbin Jiang, Zhen Wang, KaiYing Wu, Wen Zhang, Fei Wen

Publication date 2026/5/3

Conference ICASSP 2026-2026 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)

Pages 16177-16181

Publisher IEEE

Description Generative speech enhancement (SE) methods, typically implemented with diffusion or flow matching, exhibit strong generalization to unseen acoustic conditions. However, they often underperform discriminative approaches in matched, in-domain settings. Recent attempts to close this gap have relied on pretrained discriminative models or cascaded multi-flow architectures, which add inference steps and increase computational cost. To address these limitations, we propose HyFlowSE, a flow matching SE framework with hybrid generative–discriminative learning. HyFlowSE leverages neural ordinary differential equations (ODEs) for end-to-end training and jointly optimizes generative and discriminative objectives within a single model. Experiments on popular benchmark datasets show that HyFlowSE, with only 5.2 M parameters, outperforms other generative SE methods across nearly all evaluation metrics, with ...

Total citations Cited by 2



6. 论文出版信息

Authors

Figures

References

Keywords

Metrics

Footnotes

Published in: ICASSP 2026 - 2026 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)

Date of Conference: 03-08 May 2026

DOI: 10.1109/ICASSP55912.2026.11464634

Date Added to IEEE Xplore: 21 April 2026

Publisher: IEEE

ISBN Information:

Electronic ISBN:979-8-3315-6701-9

Print on Demand(PoD) ISBN:979-8-3315-6702-6

ISSN Information:

Electronic ISSN: 2379-190X

Print on Demand(PoD) ISSN: 1520-6149

Funding Agency:

10.13039/100002465-Delta

10.13039/501100013254-National College Students Innovation and Entrepreneurship Training Program

10.13039/501100001809-National Natural Science Foundation of China

Published: 2019

Show More

IEEE DiscoveryPoint for Semiconductors

Introducing IEEE DiscoveryPoint for Semiconductors: The All-in-One Platform for Engineers

Access Platform

IEEE

7. 论文正文

HYFLOWSE: HYBRID END-TO-END FLOW-MATCHING SPEECH ENHANCEMENT VIA GENERATIVE-DISCRIMINATIVE LEARNING

Yang Zhang¹, Wenbin Jiang^{1,*}, Zhen Wang¹, KaiYing Wu¹, Wen Zhang¹, Fei Wen²

¹ School of Communication Engineering, Hangzhou Dianzi University, Hangzhou, China

² School of Information Science and Electronic Engineering, Shanghai Jiao Tong University, Shanghai, China

ABSTRACT

Generative speech enhancement (SE) methods, typically implemented with diffusion or flow matching, exhibit strong generalization to unseen acoustic conditions. However, they often underperform discriminative approaches in matched, in-domain settings. Recent attempts to close this gap have relied on pretrained discriminative models or cascaded multi-flow architectures, which add inference steps and increase computational cost. To address these limitations, we propose HyFlowSE, a flow matching SE framework with hybrid generative-discriminative learning. HyFlowSE leverages neural ordinary differential equations (ODEs) for end-to-end training and jointly optimizes generative and discriminative objectives within a single model. Experiments on popular benchmark datasets show that HyFlowSE, with only 5.2 M parameters, outperforms other generative SE methods across nearly all evaluation metrics, with especially pronounced gains at low signal-to-noise ratios.

Index Terms— speech enhancement, flow matching, hybrid learning, generative model, diffusion

1. INTRODUCTION

Speech enhancement (SE) [1, 2], a core technology in modern communication, has increasingly shifted from traditional methods [3, 4] to deep learning-based approaches. For years, the field has been dominated by discriminative models [5, 6], which learn a mapping from noisy to clean speech or predict a mask to recover the clean speech. However, these models often generalize poorly to unseen noise types and yield overly-smoothed speech with artifacts.

To overcome these limitations, generative models have become a powerful alternative. Early diffusion frameworks such as SGMSE [7] learn a clean speech prior but require a slow, iterative sampling process with high numbers of function evaluations (NFE). Flow Matching (FM) [8] was introduced to reduce this latency, and models like FlowSE

[9] match the performance of a 60-NFE diffusion model using only five NFEs [10]. However, FM entails a performance-efficiency trade-off in lightweight models: for example, reducing FlowSE's parameter count lowers its PESQ score [11] from 3.12 to 2.98 [9, 12].

To solve these issues with generative models, recent work has proposed two-stage or cascaded solutions. One prominent approach, StoRM [13], cascades a predictive model with a generative diffusion model for refinement. Another strategy, the cascaded two-flows method (CasFlowSE) [12], mitigates the lightweight FM problem. While CasFlowSE can raise the PESQ score of the lightweight model back to 3.05, it does so by introducing an additional estimation step. This reliance on multi-stage pipelines, whether requiring separate models or adding intermediate inference passes, fundamentally negates the primary advantage of high-efficiency generative modeling and creates a need for a single-pass, end-to-end solution.

Inspired by investigations into diverse training objectives [14], we propose a training paradigm that integrates generative and discriminative objectives. We argue that a purely generative Conditional Flow Matching Loss [8] is insufficient for restoring details in lightweight, low-NFE models. Therefore, we introduce a discriminative loss that integrates STFT and L1 objectives for precise supervision. Our proposed method, HyFlowSE, implements this hybrid paradigm to successfully resolve the performance trade-off in lightweight models. The main contributions of this paper are summarized as follows: 1) We propose HyFlowSE, a novel end-to-end framework that integrates a discriminative objective into Flow Matching training, effectively resolving the performance degradation of lightweight models at low NFEs without adding inference overhead. 2) We demonstrate that HyFlowSE achieves state-of-the-art (SOTA) performance among comparable FM models, with strong results in challenging low-SNR environments.

2. PRELIMINARIES

2.1. Flow Matching-Based Speech Enhancement

Flow Matching is an emerging generative modeling framework that defines a continuous, invertible transformation from a simple prior distribution to a complex data distribution by

*This work was supported by the Yangtze River Delta Science and Technology Innovation Community Joint Research Project 2024CSJGG1100, the National College Students Innovation and Entrepreneurship Training Program of China (No. 202510336068), and the National Natural Science Foundation of China (NSFC) under Grant 62271314. The corresponding author is Wenbin Jiang (wbjiang@hdu.edu.cn).

modeling an Ordinary Differential Equation (ODE). Due to its deterministic nature, flow matching often requires very few inference steps (NFE) to achieve satisfactory performance[8]. A Continuous Normalizing Flow (CNF) [15] is defined by an ODE:

$$\frac{d\psi_t(x)}{dt} = v_t(\psi_t(x)) \quad (1)$$

where $\psi_t(x)$ defines the Flow from an initial state $x_1 \sim p_1$ to a target state $x_0 \sim p_0$, and $v_t(\cdot)$ is the Vector Field defining the dynamics of this flow. The goal of speech enhancement is to learn a vector field $v_t(x_t|y)$ conditioned on the noisy speech y . Learning this vector field directly is difficult. Conditional Flow Matching (CFM) significantly simplifies the training objective by introducing a probability path $p_t(x_t|x_0, y)$ and its corresponding target vector field $v_t(x_t|x_0, y)$, both conditioned on x_0 and y . FlowSE [9] employs a modified optimal transport (OT) path:

$$\mu_t(x_0, y) = (1-t)x_0 + ty, \quad (2)$$

$$\sigma_t = t\sigma \quad (3)$$

The corresponding target vector field is:

$$v_t(x_t|x_0, y) = \frac{d\sigma_t/dt}{\sigma_t}(x_t - \mu_t(x_0, y)) + \frac{d\mu_t(x_0, y)}{dt} \quad (4)$$

FlowSE trains a vector field model $v_\theta(x_t, y, t)$ to match the above target using the CFM loss:

$$L_{\text{CFM}} = \mathbb{E} \|v_\theta(x_t, y, t) - v_t(x_t|x_0, y)\|^2 \quad (5)$$

During inference, one simply samples x_1 from the prior distribution $p_1(x_1|y) = \mathcal{N}(y, \sigma^2 \mathbf{I})$ and then uses a numerical integrator like Euler's method to solve the learned ODE backwards from $t = 1$ to $t = 0$ to generate enhanced speech in a few steps. The success of FlowSE demonstrates the great potential of flow matching for efficient speech enhancement. However, when lightweight network architectures are adopted to pursue ultimate deployment efficiency, its performance degrades significantly. This efficiency-performance trade-off presents a new challenge for the practical application of flow matching models.

2.2. CasFlowSE: Cascaded Flow Matching SE

To alleviate the performance degradation of lightweight flow matching models, Lee et al. [12] proposed the Cascaded Two Flows for Speech Enhancement (CasFlowSE) model. The core idea of CasFlowSE is to cascade the same flow matching model twice for coarse-to-fine enhancement. The first flow is identical to FlowSE, and its output $D_\theta(x_1, y)$ is a rough estimate of the clean speech. This rough estimate is then used as both a conditioning variable and the starting point for the second flow, effectively setting the prior distribution to $\mathcal{N}(D_\theta(x_1, y), \sigma^2 \mathbf{I})$. The second flow performs further refinement based on this to produce the final high-quality speech.

The loss function for CasFlowSE is a weighted sum of multiple CFM losses:

$$L_{\text{total}} = \lambda_1 L_{\text{CFM1}} + \lambda_2 L_{\text{CFM2}} + \lambda_3 L_{\text{CFM3}} \quad (6)$$

where L_{CFM1} and L_{CFM2} correspond to the training objectives of the two flows, respectively, and L_{CFM3} is an additional constraint designed to force the output of the first flow to be closer to the real clean speech. Experiments show that CasFlowSE effectively improves the performance of lightweight models. Nevertheless, this performance improvement comes at the cost of inference efficiency. The total NFE becomes the sum of the steps required by both flows. In essence, this represents a computation-for-performance compromise that deviates from the original intention of flow matching to pursue highly efficient inference.

3. PROPOSED HYFLOWSE METHOD

We propose HyFlowSE, an end-to-end hybrid learning framework integrating generative and discriminative objectives. As illustrated in Fig. 1, the core idea is to enhance lightweight Flow Matching models using a hybrid loss. This is made possible by a differentiable ODE solver, which unifies both objectives in a single training pass for end-to-end optimization without adding inference overhead.

3.1. Neural ODE

The key mechanism enabling this end-to-end optimization is the concept of a Neural Ordinary Differential Equation (Neural ODE) [15]. A Neural ODE defines continuous dynamics by parameterizing the derivative of a hidden state $z(t)$ with a neural network f_θ :

$$\frac{dz(t)}{dt} = f_\theta(z(t), t) \quad (7)$$

where $z(t)$ is the hidden state at a continuous time t , and the network f_θ with parameters θ learns its instantaneous rate of change.

By modeling the generation process as a continuous-time dynamic system, the entire trajectory becomes differentiable with respect to the model parameters. This differentiability serves as the technical foundation of HyFlowSE, enabling the seamless integration of generative and discriminative objectives. Consequently, the gradients can be propagated through the ODE solver to optimize the vector field network in a strictly end-to-end manner.

3.2. Hybrid Generative-Discriminative Objective

The total learning objective of HyFlowSE is a weighted sum of the standard conditional flow matching loss, L_{CFM} , and our proposed discriminative loss, L_{DISC} :

$$L_{\text{HyFlowSE}} = \alpha L_{\text{CFM}} + \beta L_{\text{DISC}} \quad (8)$$

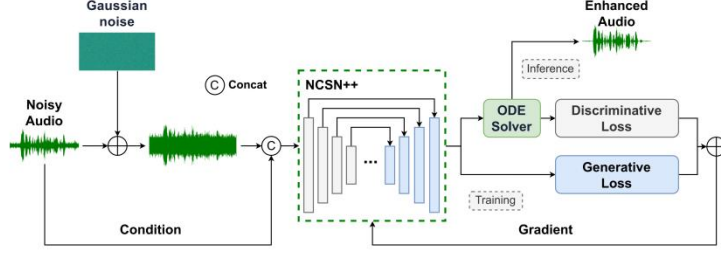


Fig. 1. An illustration of the HyFlowSE framework. It computes the generative loss and the discriminative loss in parallel.

where α and β are hyperparameters that balance the two objectives.

As the generative objective, L_{CFM} is consistent with the definition in FlowSE [9]. It learns the continuous transformation of the data distribution by matching the target vector field. Its formula is:

$$L_{CFM} = \mathbb{E}_{t, x_0, y} \|v_\theta(x_t, y, t) - v_t(x_t | x_0, y)\|^2 \quad (9)$$

On the other hand, the discriminative objective L_{DISC} provides a more direct and fine-grained supervisory signal. It directly computes the difference between the enhanced speech $\hat{x}_0$, generated by the complete ODE integration from $t = 1$ to $t = 0$, compared against the ground-truth clean speech x_0 . This loss fuses metrics from both the time and frequency domains:

$$L_{DISC} = w_1 L_{L1} + w_2 L_{SC} + w_3 L_{MAG} \quad (10)$$

where $L_{L1} = \mathbb{E} \|\hat{x}_0 - x_0\|_1$ is the time-domain L1 loss, which ensures waveform-level accuracy. To simultaneously optimize speech quality from a frequency-domain perspective that is more consistent with human auditory perception, we employ multi-resolution STFT losses following [16], consisting of the Spectral Convergence loss defined as $L_{SC} = \frac{\| |\text{STFT}(\hat{x}_0)| - |\text{STFT}(x_0)| \|_F}{\| |\text{STFT}(x_0)| \|_F}$ and the Log STFT Magnitude loss defined as $L_{MAG} = \frac{1}{N} \|\log |\text{STFT}(\hat{x}_0)| - \log |\text{STFT}(x_0)|\|_1$, where $\|\cdot\|_F$ and $\|\cdot\|_1$ denote the Frobenius norm and the L_1 norm, respectively, and N denotes the number of elements in the magnitude spectrogram.

3.3. End-to-End Training via Differentiable ODE Solver

Our hybrid objective is trained end-to-end using a differentiable ODE solver. We implement this using the `odeint` function from the `torchdiffeq` library [17]. In each training iteration, we compute the standard L_{CFM} while simultaneously using `odeint` to perform a full NFE-step integration to generate $\hat{x}_0$ for the L_{DISC} calculation. The gradients from both losses are then combined to update the vector field network v_θ . This differentiable pipeline is the technical core of HyFlowSE, enabling the discriminative objective to effectively guide the entire generative process.

4. EXPERIMENTS

4.1. Datasets

Our experiments are conducted on the VoiceBank+DEMAND dataset [18], which consists of clean speech from VCTK[19] mixed with noise from DEMAND[20]. Additionally, we use three datasets based on the WSJ0 corpus [21], generated using the open-source code from StoRM [13]: WSJ0+CHiME3 [22], which includes a high-SNR (H, [0, 20] dB) version and a low-SNR (L, [-4, 16] dB) version, and WSJ0+Reverb for evaluating performance in reverberant conditions.

4.2. Implementation Details

Our models are based on the NCSN++ architecture. We use four configurations of different scales by adjusting hyperparameters, resulting in models with 65.6M, 27.8M, 11.7M, and 5.2M parameters. The proposed models are named HyFlowSE, HyFlowSE(M), HyFlowSE(S), and HyFlowSE(T), corresponding to these four parameter configurations, respectively. We compare them against representative Flow Matching baselines, including the 65.6M FlowSE, the 27.8M FlowSE(M)[9], and CasFlowSE [12]. For a fair comparison, baseline results are either cited from the original papers or, for unavailable NFE=5 settings, reproduced by us using the official open-source code.

All models are trained with the Adam [23] optimizer with a learning rate of 1×10^{-4} . The key weights for our hybrid loss are $\alpha = 2 \times 10^{-4}$ for the generative term, and $(w_{L1}, w_{SC}, w_{MAG}) = (1.0, 0.5, 0.5)$ for the discriminative term.

Our evaluation proceeds in two stages. First, to demonstrate the framework's effectiveness and scalability, we conduct a performance comparison of all four HyFlowSE variants on the VoiceBank+DEMAND dataset, focusing on the impact of different model capacities. Second, we benchmark our smallest 5.2M model (HyFlowSE(T)) against larger competitors on the WSJ0 datasets to validate its competitive performance, which is our core hypothesis. For all comparisons, the number of function evaluations (NFE) is fixed at 5. Performance is evaluated using standard metrics including

Table 1. Speech enhancement performances for proposed and compared methods on the VoiceBank+DEMAND datasets.

Trained and Tested on VoiceBank+DEMAND									
METHOD	Para.(M)	PESQ	eSTOI	SI-SDR	WVMOS	DNSMOS	SIG	BAK	OVRL
FlowSE	65.6	3.12	0.88	18.95	4.34	3.58	3.49	4.05	3.21
FlowSE(M)	27.8	2.98	0.87	18.97	4.30	3.58	3.48	4.05	3.20
CasFlowSE	27.8	3.05	0.88	19.13	4.26	3.58	3.48	4.03	3.20
HyFlowSE	65.6	3.28	0.89	19.12	4.43	3.52	3.55	4.05	3.26
HyFlowSE(M)	27.8	3.26	0.89	19.20	4.42	3.52	3.54	4.05	3.26
HyFlowSE(S)	11.7	3.25	0.88	19.20	4.42	3.52	3.54	4.05	3.25
HyFlowSE(T)	5.2	3.21	0.88	19.09	4.39	3.52	3.54	4.05	3.26

Table 2. Speech enhancement performances for proposed and compared methods on the WSJ0 series datasets.

Trained and Tested on WSJ0+CHiME3 (H)					
METHOD	Para.(M)	PESQ	eSTOI	SI-SDR	DNSMOS
FlowSE(M)	27.8	3.00	0.93	18.70	4.05
CasFlowSE	27.8	3.15	0.94	19.84	4.05
HyFlowSE(T)	5.2	3.31	0.95	19.96	3.90

Trained and Tested on WSJ0+CHiME3 (L)					
METHOD	Para.(M)	PESQ	eSTOI	SI-SDR	DNSMOS
FlowSE(M)	27.8	2.64	0.89	15.34	4.01
CasFlowSE	27.8	2.64	0.89	15.96	4.01
HyFlowSE(T)	5.2	3.09	0.93	17.39	3.88

Trained and Tested on WSJ0+Reverb					
METHOD	Para.(M)	PESQ	eSTOI	SI-SDR	DNSMOS
FlowSE(M)	27.8	2.45	0.84	4.43	3.96
CasFlowSE	27.8	2.80	0.89	7.90	3.97
HyFlowSE(T)	5.2	2.95	0.89	2.80	3.91

Para.(M), PESQ scores [11], DNSMOS [24], wav2vec MOS (WVMOS) [25], eSTOI [26], DNSMOS P.835 [27] including SIG, BAK, OVRL, and Scale-Invariant Signal-to-Distortion Ratio (SI-SDR) [28].

4.3. Results

4.3.1. Evaluation results on VoiceBank+DEMAND dataset

We first evaluate the performance of the compared methods on the VoiceBank+DEMAND dataset. Audio samples and additional materials are available online¹. As shown in Table 1, our proposed hybrid paradigm demonstrates superior performance on the VoiceBank+DEMAND dataset. On one hand, our method achieves substantial gains in objective metrics measuring signal fidelity and intelligibility. Specifically, with a comparable parameter count of 27.8M, HyFlowSE(M) secures a distinctly higher PESQ score than both FlowSE(M) and CasFlowSE. This advantage is further underscored by parameter efficiency, where the compact HyFlowSE(T) (5.2M) surpasses the significantly larger FlowSE (65.6M) in terms of PESQ. Moreover, the method exhibits remarkable robust-

ness to lightweighting; despite a drastic parameter reduction from 65.6M in HyFlowSE to 5.2M in HyFlowSE(T), the model sustains a performance level comparable to the full-size model with a stable SI-SDR. On the other hand, while baseline models hold a marginal edge in the non-intrusive DNSMOS metric, our hybrid design prioritizes and excels in signal-level fidelity.

4.3.2. Evaluation results on WSJ0 dataset

We further evaluate the performance on the more challenging WSJ0 dataset, and the results are presented in Table 2. The data demonstrates that the superiority of our method is particularly evident in adverse acoustic conditions. In the low signal-to-noise ratio scenario of WSJ0+CHiME3 (L), the compact HyFlowSE(T) with only 5.2M parameters achieves a substantial improvement in PESQ, significantly surpassing both 27.8M baseline models. Similarly, in the reverberant setting, our method outperforms FlowSE(M) and CasFlowSE in terms of perceptual quality. A notable trade-off appears in the reverberation task where CasFlowSE attains a higher SI-SDR, yet our HyFlowSE(T) delivers superior PESQ performance. Despite a slight advantage for the baselines in DNSMOS across all sub-datasets, HyFlowSE(T) consistently leads in key metrics for speech quality and intelligibility, such as PESQ and eSTOI, underscoring its effectiveness in producing high-fidelity speech under challenging environments.

5. CONCLUSION

To address the performance bottleneck of lightweight Flow Matching models at low NFE, we proposed HyFlowSE, a hybrid training framework that enhances a model's intrinsic capabilities without any inference overhead. Using differentiable neural ODEs, HyFlowSE deeply fuses a generative CFM loss with a powerful discriminative loss, enabling stable end-to-end optimization during training. Experimental results confirm that our 5.2M HyFlowSE(T) model achieves SOTA performance and significantly outperforms baselines several times its size, particularly in challenging low-SNR scenarios with an NFE of just 5. The HyFlowSE framework provides a new path to break the trade-off between efficiency and performance in the enhancement of generative speech.

¹<https://zhangyang77.github.io/HyFlowSE/>

6. REFERENCES

- [1] Philipos C Loizou, *Speech enhancement: theory and practice*, CRC press, 2007.
- [2] Richard C Hendriks, Timo Gerkmann, and Jesper Jensen, *DFT-domain Based Single-microphone Noise Reduction for Speech Enhancement: A Survey of the State-of-the-art*, vol. 11, Morgan & Claypool Publishers, 2013.
- [3] Yariv Ephraim and David Malah, "Speech enhancement using a minimum-mean square error short-time spectral amplitude estimator," *IEEE Transactions on acoustics, speech, and signal processing*, vol. 32, no. 6, pp. 1109–1121, 2003.
- [4] Timo Gerkmann and Emmanuel Vincent, "Spectral masking and filtering," *Audio source separation and speech enhancement*, pp. 65–85, 2018.
- [5] Yanxin Hu, Yun Liu, Shubo Lv, et al., "DCCRN: Deep Complex Convolution Recurrent Network for Phase-Aware Speech Enhancement," in *Proc. ISCA Interspeech*, 2020, pp. 2472–2476.
- [6] Ke Tan and DeLiang Wang, "Learning complex spectral mapping with gated convolutional recurrent networks for monaural speech enhancement," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 28, pp. 380–390, 2020.
- [7] Simon Welker, Julius Richter, and Timo Gerkmann, "Speech enhancement with score-based generative models in the complex STFT domain," in *Proc. Interspeech 2022*, 2022, pp. 2928–2932.
- [8] Yaron Lipman, Ricky TQ Chen, Heli Ben-Hamu, Maximilian Nickel, and Matt Le, "Flow matching for generative modeling," *arXiv preprint arXiv:2210.02747*, 2022.
- [9] Seonggyu Lee, Seon Cheong, Sangwook Han, and Jong Won Shin, "Flowse: Flow matching-based speech enhancement," in *ICASSP 2025 - 2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2025.
- [10] Bunlong Lay, Simon Welker, Julius Richter, and Timo Gerkmann, "Reducing the prior mismatch of stochastic differential equations for diffusion-based speech enhancement," *arXiv preprint arXiv:2302.14748*, 2023.
- [11] ITU-T Recommendation, "Perceptual Evaluation of Speech Quality (PESQ): An objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs," *Rec. ITU-T P. 862*, 2001.
- [12] Seonggyu Lee, Seon Cheong, Sangwook Han, Kihyuk Kim, and Jong Won Shi, "Speech enhancement based on cascaded two flow," *arXiv preprint arXiv:2508.06842*, 2025.
- [13] Jean-Marie Lemerrier, Julius Richter, Simon Welker, and Timo Gerkmann, "Storm: A diffusion-based stochastic regeneration model for speech enhancement and dereverberation," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 31, pp. 2724–2737, 2023.
- [14] Julius Richter, Danilo De Oliveira, and Timo Gerkmann, "Investigating training objectives for generative speech enhancement," in *ICASSP 2025 - 2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2025, pp. 1–5.
- [15] Ricky TQ Chen, Yulia Rubanova, Jesse Bettencourt, and David K Duvenaud, "Neural ordinary differential equations," *Advances in neural information processing systems*, vol. 31, 2018.
- [16] Ryuichi Yamamoto, Eunwoo Song, and Jae-Min Kim, "Parallel wavegan: A fast waveform generation model based on generative adversarial networks with multi-resolution spectrogram," in *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, IEEE, 2020, pp. 6199–6203.
- [17] Ricky T. Q. Chen, "torchdiffeq," 2018.
- [18] Cassia Valentini Botinhao, Xin Wang, Shinji Takaki, and Junichi Yamagishi, "Investigating rnn-based speech enhancement methods for noise-robust text-to-speech," in *9th ISCA speech synthesis workshop*, 2016, pp. 159–165.
- [19] Christophe Veaux, Junichi Yamagishi, and Simon King, "The voice bank corpus: Design, collection and data analysis of a large regional accent speech database," in *Oriental International Conference on Speech Database and Assessments CO-COSDA*, IEEE, 2013, pp. 1–4.
- [20] Joachim Thiemann, Nobutaka Ito, and Emmanuel Vincent, "The diverse environments multi-channel acoustic noise database (demand): A database of multichannel environmental noise recordings," in *Proceedings of Meetings on Acoustics*, Acoustical Society of America, 2013, vol. 19, p. 035081.
- [21] John S Garofolo, David Graff, Doug Paul, and David Pallett, "Csr-i (wsj0) complete," (*No Title*), 2007.
- [22] Jon Barker, Ricard Marxer, Emmanuel Vincent, and Shinji Watanabe, "The third 'chime' speech separation and recognition challenge: Dataset, task and baselines," in *2015 IEEE workshop on automatic speech recognition and understanding (ASRU)*, IEEE, 2015, pp. 504–511.
- [23] Diederik P Kingma and Jimmy Ba, "Adam: A method for stochastic optimization," in *Int. Conf. Learn. Represent.*, 2014, pp. 1–15.
- [24] Oscar Figuerola Salas, Velibor Adzic, and Hari Kalva, "Subjective quality evaluations using crowdsourcing," in *2013 Picture Coding Symposium (PCS)*, IEEE, 2013, pp. 418–421.
- [25] Pavel Andreev, Aibek Alanov, Oleg Ivanov, and Dmitry Vetrov, "Hifi++: a unified framework for bandwidth extension and speech enhancement," in *ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, IEEE, 2023, pp. 1–5.
- [26] Jesper Jensen and Cees H Taal, "An algorithm for predicting the intelligibility of speech masked by modulated noise maskers," *IEEE/ACM Trans. ASLP*, vol. 24, no. 11, pp. 2009–2022, 2016.
- [27] Chandan KA Reddy, Vishak Gopal, and Ross Cutler, "Dnsmos p. 835: A non-intrusive perceptual objective speech quality metric to evaluate noise suppressors," in *ICASSP 2022-2022 IEEE international conference on acoustics, speech and signal processing (ICASSP)*, IEEE, 2022, pp. 886–890.
- [28] Jonathan Le Roux, Scott Wisdom, Hakan Erdogan, et al., "SDR-half-baked or well done?," in *Proc. IEEE ICASSP*, IEEE, 2019, pp. 626–630.

三、科研积分—科研项目—国家级大创结题—负责人（5分）

通信工程学院大创项目核查清单								
查询人学号	查询人姓名	项目名称	项目编号	项目负责人	项目成员	指导教师	项目结题情况	备注
23080440	李京瑞	面向通信网络主链故障多源信息检测与定位方法研究	202510336068	李京瑞	周海、王强、贾均耀、李远臻	江文斌	已结题	
23080309	张杨	一种基于条件流匹配和声码器的语音增强算法	202510336068	张杨	吴开颖、张雯、徐为宇、李品桐	江文斌	已结题	
23080509	杨语歌	面向脑电图模式识别的ADHD神经网络构建方法研究	202510336048	宋妍慧	贾均耀、廖耀坤、杨语歌、李远臻	吴瑞坡	已结题	
23080408	廖丹丹	应对老龄化亚健康健康问题关键技术——基于5G通信的移动监护仪	S202410336147	管静宜	廖丹丹、荆凯	杨国伟	立项未结题	
23080408	廖丹丹	ATIS与雷达数据融合的船舶轨迹异常检测系统	S202510336010	廖丹丹	蒋伟洁、金文理、周润、王宇桐	孙超、刘泰山	立项未结题	
23080422	夏敬博	车联网场景下多RIS辅助定位	S202510336015	夏敬博	俞晨帆、史婉梅、陈鑫雷、郑晨曦	曾嵘	已结题	
23080407	胡远凝	融合智测—智能便携式多维生命体征监测系统	202510336052	荆凯	胡远凝、管静宜、蒋惠霞、方荣品	杨国伟、周雪芳	立项未结题	
23080717	王梓豪	多源信息融合的海上目标检测识别系统	S202510336024	卢曦	张中源、王梓豪、王祺、魏欣萌	孙超	已结题	
23080717	王梓豪	基于动态最大似然的全息MIMO定位方法研究	S202410336073	解明子	郑思晗、姚叶哲鸣、李翔岩、王梓豪	胡安中、刘天乐	已结题	
23080540	俞涵韬	基于物理约束Transformer的三维智慧气象—电磁特性一体化预报系统	202510336034	詹锦瑜	陈梓毓、吕婷婷、俞涵韬、董金强	侯壮玉、应娜	已结题	
23080407	胡远凝	基于心脑血管慢性病预防监测关键技术——多维生命体征监测系统研发	S202410336100	胡远凝	卢曦、蒋惠霞、朱欣烨、张世丞	周雪芳、杨国伟	立项未结题	
23080807	詹锦瑜	基于物理约束Transformer的三维智慧气象—电磁特性一体化预报系统	202510336034	詹锦瑜	陈梓毓、吕婷婷、俞涵韬、董金强	侯壮玉、应娜	已结题	
23080305	王一淼	基于城适应知识蒸馏的多模态水冰分类研究	S202510336064	陈坤	王一淼、刘晓敏、郭子璇	毕美华、张永强	已结题	
23080430	陈坤	基于城适应知识蒸馏的多模态水冰分类研究	S202510336064	陈坤	王一淼、刘晓敏、郭子璇	毕美华、张永强	已结题	
23080201	吕婷婷	基于物理约束Transformer的三维智慧气象—电磁特性一体化预报系统	202510336034	詹锦瑜	陈梓毓、吕婷婷、俞涵韬、董金强	侯壮玉、应娜	已结题	
23080211	董金强	基于物理约束Transformer的三维智慧气象—电磁特性一体化预报系统	202510336034	詹锦瑜	陈梓毓、吕婷婷、俞涵韬、董金强	侯壮玉、应娜	已结题	
23080803	卢曦	基于心脑血管慢性病预防监测关键技术——多维生命体征监测系统研发	S202410336100	胡远凝	卢曦、蒋惠霞、朱欣烨、张世丞	周雪芳、杨国伟	已结题	
23150522	谢廷耀	智慧先锋——基于智能视觉的施工区实时防撞预警系统	S202510336117	柯昊天	谢廷耀、王乐源、孔梓凯、黄晓晴	史晓颖、陈婧	立项未结题	
23080803	卢曦	多源信息融合的海上目标监测系统	S202510336024	卢曦	张中源、王梓豪、王祺、魏欣萌	孙超、刘泰山	已结题	
23080607	宋妍慧	面向脑电图模式识别的ADHD神经网络构建方法研究	202510336048	宋妍慧	贾均耀、廖耀坤、杨语歌、李远臻	吴瑞坡	已结题	
23080637	李远臻	面向脑电图模式识别的ADHD神经网络构建方法研究	202510336048	宋妍慧	贾均耀、廖耀坤、杨语歌、李远臻	吴瑞坡	已结题	
23080615	廖耀坤	面向脑电图模式识别的ADHD神经网络构建方法研究	202510336048	宋妍慧	贾均耀、廖耀坤、杨语歌、李远臻	吴瑞坡	已结题	

四、专家推荐信

➤ 殷海兵教授

专家推荐信

本人应张杨同学请求，结合对其本科阶段学习和科研表现的了解，现就该生的学习能力、科研素养及培养潜力作如下推荐。

张杨同学具有较好的专业基础和较强的科研兴趣。在本科阶段的学习和科研训练过程中，从最初的专业知识积累，到后来能够较为独立地开展科研工作，她表现出了较好的学习能力、自我驱动力和科研成长性。

在科研方面，张杨同学主要从事生成式语音增强研究。经过持续训练，她已经取得了一系列阶段性成果。她以第一作者身份完成的 HyFlowSE 被 ICASSP 2026 录用，同时还有其他第一作者研究工作，并参与完成了 Interspeech 2026 相关论文和语音增强相关发明专利。这些经历表明，她已经接受了较为系统的本科科研训练，并具备一定的独立开展研究工作的能力。

张杨同学对科研具有较强的兴趣和持续投入的能力。在实验过程中遇到困难时，她能够认真分析并尝试不同的解决方案。在论文撰写、实验修改以及研究方案完善过程中，她表现出了较好的耐心和自我反思能力。

同时，张杨同学具有较好的团队合作意识和项目执行能力。她曾负责国家级大学生创新创业训练计划项目及院级大学生创新创业训练计划项目。在项目推进过程中，她能够与团队成员进行有效沟通，合理推进研究任务，并在实践中不断提升自己的组织协调能力。

总体来看，张杨同学学习能力较强，科研兴趣明确，具有较好的专业基础、科研素养和团队协作能力。本科阶段持续的科研训练使她逐渐形成了较为成熟的科研意识，并表现出了良好的学术发展潜力和培养前景。本人认为该生综合表现突出，具有进一步深入开展科研工作的良好基础，故予以推荐。

推荐人签字：

日期：2026 年 8 月 30 日

➤ 赵宜楠教授

专家推荐信

本人应张杨同学请求,根据对其本科阶段学习、科研及创新实践情况的了解,现就该生的科研素养、创新能力和发展潜力作如下推荐。

张杨同学在本科期间表现出了较强的学习能力和科研主动性。她在完成专业课程学习的同时,能够主动拓展课堂之外的知识。从专业基础知识的学习,到逐步开展文献调研、模型设计、实验分析和论文撰写,她在科研训练过程中表现出了较快的成长速度和较强的执行能力。

张杨同学能够围绕较为明确的研究问题持续开展探索。她长期关注生成式方法在语音增强任务中的应用。相关工作已经形成多项成果,其中以第一作者身份完成的 HyFlowSE 被 ICASSP 2026 录用为 Oral,同时还有第一作者研究工作处于评审阶段,并参与完成了 Interspeech 2026 相关论文。

除论文研究外,张杨同学还积极参与各类科研创新项目。她曾负责国家级大学生创新创业训练计划项目并完成结题,负责院级大学生创新创业训练计划项目并获得优秀结题。通过这些项目,她不仅进一步锻炼了科研能力,也积累了项目规划、任务推进、团队协作以及沟通协调等方面的经验。

在科研实践中,张杨同学具有较好的主动思考意识。面对具体研究任务,她能够主动查阅资料、分析已有方法的特点与不足,并结合实验结果不断调整研究思路。这种发现问题、分析问题并持续改进的能力,是其科研成长过程中较为突出的特点。

总体而言,张杨同学具有较强的科研主动性、创新意识和项目执行能力,本科阶段已经取得了较为丰富的科研成果,并形成了一定的独立研究能力。本人认为她具有良好的学术发展潜力和进一步培养价值,故予以推荐。

推荐人签字: 

日期: 2026 年 8 月 30 日

➤ 夏永祥教授

专家推荐信

本人应张杨同学请求，结合对其本科阶段科研工作和学术表现的了解，现就该生的专业基础、科研能力及学术潜力作如下推荐。

张杨同学在校期间学习认真，专业基础扎实，对科研工作具有较为浓厚的兴趣。本科阶段，她较早参与科研训练，主要围绕生成式模型在语音增强任务中的应用。经过持续的科研训练，她逐步形成了较为明确的研究方向，并具备了较好的文献阅读、问题分析、实验设计和论文撰写能力。

在科研方面，张杨同学表现较为突出。她围绕生成式语音增强开展了多项连续性的研究工作。其以第一作者身份完成的论文《HyFlowSE: Hybrid End-To-End Flow-Matching Speech Enhancement via Generative-Discriminative Learning》被语音领域顶会 ICASSP 2026 录用为 Oral。此外，她还有相关第一作者论文工作，并作为共同作者参与完成了 Interspeech 2026 论文，同时参与申请了语音增强相关发明专利。

在科研过程中，张杨同学并不满足于简单复现已有方法，而是能够主动阅读相关领域文献，并尝试提出新的解决思路。在实验结果不理想、研究方案需要调整以及论文修改过程中，她能够认真分析原因并持续完善工作，体现出较好的科研耐心、独立思考能力和钻研精神。

在我看来，张杨同学在本科阶段已经形成了较为连续的科研积累，并表现出了良好的科研能力和创新意识。她能够围绕具体研究问题持续深入，在模型设计、实验分析等方面均积累了较为扎实的经验，具有进一步培养的良好基础。

综合来看，张杨同学专业基础扎实，科研能力较强，在本科阶段已经表现出较好的学术素养和发展潜力。本人认为该生具有进一步开展深入研究的能力和潜质，故予以推荐。

推荐人签字：

日期：2026 年 8 月 30 日