

附件 1

南开大学推荐免试攻读研究生申请表暨诚信承诺书

学号	2310137	姓名	武宇航	性别	男
民族	汉	出生日期	2005 年 5 月 25 日	政治面貌	共青团员
学院	密码与网络空间安全学院		专业	信息安全、法学	
手机号码	13111342470		电子邮箱	2310137@mail.nankai.edu.cn	
申请类型 (单选)	☒ 普通类推免名额 ☐ 支教团推免名额				
本人与学院遴选指标相关的材料列表	(学生可将本人与本学院遴选指标相关的材料列在下方，如科研成果、竞赛获奖、科研训练表现、参军入伍服兵役、志愿服务、国际组织实习等) 1. 全国大学生信息安全竞赛，一等奖 2. WISE 2026 第一作者论文，CCF-B 类 3. ICIC 2026 第一作者论文，CCF-C 类 4. 志愿服务时长共 79 小时				
本人充分了解并理解《南开大学推荐优秀应届本科毕业生免试攻读研究生工作实施办法》(南党发〔2026〕37 号)和所在学院《推免细则》，并郑重承诺： 1. 本人所提供的一切材料真实可靠，无弄虚作假，如有不实之处，本人愿意承担由此造成的一切后果； 2. 如获得推免生资格，本人将慎重填报志愿，绝不浪费学校推免名额。 承诺人： 年 月 日					

(本申请书必须由申请人亲笔签名方为有效，交所在学院本科教学工作办公室备查)

综合素质评价审查材料目录（不包含公能素质部分）

学院：密码与网络空间安全学院 专业：信息安全、法学双学位班 学号：2310137 姓名：武宇航

请根据个人实际情况填写此目录（其他项目可自行添加行），并按填写顺序排序所有证明材料
后附证明材料要求真实、直观。
此目录纸张横向单面打印、附在“承诺书”后

		加分项目 ① 赛事获奖请直接直接填写右边各列 ② 论文加分请写明论文题目后填写右边各列 ③ 辅修专业课程请写明辅修专业 ④ 参军入伍及系列情况写“参军入伍” 每类事项行数不够请在该类目下自行添加	此处填写获奖等级： ① 赛事获奖请填写“一等奖/二等奖/三等奖/金奖/银奖/铜奖” ② 论文发表请填写“发表期刊-类别”（eg. “IEEE TDSC”-CCF 英文 A 类） ③ 辅修专业课程总学分数 ④ 参军入伍/嘉奖/优秀士兵/三等功及以上	证明材料类别： ① 赛事获奖证书/获奖公告页面截图 （材料里需标注出个人获奖信息所在行） ② 录用邮件页面截图/期刊网页截图+论文原文文本（证明自己是一作） ③ 辅修成绩单
特殊学术专长	重要赛事	高教社杯全国大学生数学建模竞赛本科组		
		全国大学生电子设计竞赛本科组		
		“挑战杯”全国大学生课外学术科技作品竞赛		
		中国国际大学生创新大赛（原中国国际“互联网+”大学生创新创业大赛）		
科研训练	各专业代表性赛事	ACM-ICPC 亚洲区域赛		
		全国大学生信息安全竞赛	一等奖	赛事获奖证书
		全国密码技术竞赛		
		全国大学生计算机系统能力大赛		

		全国大学生物联网设计竞赛		
	一般赛事	全国大学生数学建模竞赛天津赛区		
		天津市“挑战杯”大学生课外学术科技作品竞赛		
		中国国际大学生创新大赛天津赛区(原中国国际“互联网+”大学生创新创业大赛天津赛区)		
		天津市物联网大赛		
		京津冀大学生信息安全网络攻防大赛(原天津市大学生信息安全网络攻防大赛)		
	辅修			
特殊学术专长	论文发表	RSCA-MSRNet: Residual-Subspace Calibration and Aggregation for Fine-Grained Web Image Classification	WISE 2026-CCF 英文 B 类	CMT 系统录用页面截图、录用邮件截图、会议成功注册邮件截图、论文全文
	论文发表	LLM-Recognizer: Generative - Discriminative Feature Fusion for Sentence-level Detection	ICIC 2026-CCF 英文 C 类	会议论文集网页截图、录用邮件截图、论文全文
	论文发表			
参军入伍				
其他材料				



第十七届全国大学生信息安全竞赛 荣誉证书

编号：202401022

参赛作品： ScopeChecker-超级应用API-Scope错位检测系统

获奖等级： 一等奖

参赛高校： 南开大学

指导老师： 过辰楷

参赛队员： 裴家琪、陈宇心、刘迪乘、武宇航

教育部高等学校网络空间安全专业教学指导委员会
(秘书长单位:北京电子科技学院 代章)

二〇二四年八月



WISE 2026: Notification of paper 696

 **Microsoft CMT** <noreply@msr-cmt.org> 2026-08-07 22:18
收件人: 我 收起



发件人: Microsoft CMT noreply@msr-cmt.org

收件人: 我 2310137@mail.nankai.edu.cn

时 间: 2026-08-07 22:18:59

🚨 检测到「noreply@msr-cmt.org」不在你的通讯录中, 请注意防范风险

[添加至通讯录](#) [加入黑名单](#)

✕
不再提醒

📘 英语 ▾ 翻译为 简体中文 ▾ 翻译邮件

Dear Yuhang Wu,

We are pleased to inform you that your following paper:

696 - RSCA-MSRNet: Residual-Subspace Calibration and Aggregation for Fine-Grained Web Image Classification

has been accepted for presentation at the WISE 2026 Conference. Congratulations!

We received 782 submissions this year. Due to space limitations, WISE 2026 maintained a highly selective review process and accepted only 242 papers, resulting in an acceptance rate of around 30%. We are delighted that your paper is among the accepted submissions. You may access the reviewers' comments by logging into the CMT submission system.

To prepare and submit your camera-ready version, please carefully follow the instructions available at:

<https://conferences.sigappfr.org/wise2026/camera-ready-guidelines/>

Please also visit the conference website to complete your registration before September 2nd, 2026 (<https://conferences.sigappfr.org/wise2026/registration-fees/>). Kindly note that at least one full author registration is required for each accepted paper for it to be included in the conference proceedings and presented at the conference.

Author Console

+ Create new submission...

1 - 2 of 2 « « 1 » » Show: 25 50 100 All Clear All Filters

164		Papers	CARE_Rerank__Conflict_Aware_Evidence_Reranking_for_Trustworthy_Retrieval_Augmented_Generation__1_.pdf	Reviews	Create Camera Ready Submission
			Supplementary File Not Uploaded		
696	RSKA-MSRNet: Residual-Subspace Calibration and Aggregation for Fine-Grained Web Image Classification Hide abstract Fine-grained classification of small web images is challenging because discriminative evidence is often distributed across weak object parts, local textures, and category-specific edge patterns rather than concentrated in one salient region. Recent visual backbones improve recognition through global attention, convolutional modulation, mobile hybrid blocks, or state-space mixing, but these designs are not automatically effective when compressed into a shallow residual classifier for \$32\times32\$ images. This paper proposes \method, a compact residual classifier built around \block, a residual-subspace feature-fusion unit. Instead of inserting independent modules sequentially, \block first expands the residual channel space into multiple subspaces, then performs progressive residual aggregation among these subspaces, and finally applies lightweight channel calibration to the fused representation. This ordering is essential: channel selection is performed after cross-branch aggregation, so the gate operates on already-composed local, mid-level, and contextual cues. The training protocol, including RandAugment, Cutout, Mixup, label smoothing, staged decay, gradient clipping, and EMA evaluation, is treated as an optimisation layer rather than as a model contribution. On CIFAR-100, \block improves a matched ResNet-18 baseline from \$75.87\%\$ to \$80.15\%\$ top-1 accuracy; the complete \method reaches \$84.31\%\$ top-1 and \$96.77\%\$ top-5. Protocol-normalised comparisons with eight visual backbones and controlled ablations show that the gain comes from residual-subspace fusion rather than from merely widening the network or appending generic attention. Training dynamics, Grad-CAM, saliency, class-wise accuracy, and confusion analysis further support the same explanation.	Research Papers	RSKA_MSRNet__Residual_Subspace_Calibration_and_Aggregation_for_Fine_Grained_Web_Image_Classification (1).pdf Submission files: Supplementary File Not Uploaded Camera Ready Submission files: submission_696.zip	Accept Reviews	Camera Ready: Edit Camera Ready Submission View Camera Ready Summary

Your OpenCEMS Payment Platform order has been received!

OpenCEMS Payment Platform <sigappfr@gmail.com> (Brchber@nl1-ls11.a2hosting.com代发) 2026-08-27 19:46
收件人: 我 收起

发件人: OpenCEMS Payment Platform sigappfr@gmail.com

(Brchber@nl1-ls11.a2hosting.com代发)

收件人: 我 2310137@mail.nankai.edu.cn

时 间: 2026-08-27 19:46:07

检测到「sigappfr@gmail.com」不在你的通讯录中。请注意防范风险

添加至通讯录 加入黑名单

英语 翻译为 简体中文 翻译邮件

Thank you for your order

Hi Yuhang,

Just to let you know — we've received your order #7340, and it is now being processed.

[Order #7340] (August 27, 2026)

Product	Quantity	Price
Regular Author Registration	1	€700.00
Subtotal:		€700.00
Payment method:		Payment by credit card
Total:		€700.00
Note:		696, Yuhang Wu and Chen Chen, RSCA-MSRNet: Residual-Subspace Calibration and Aggregation for Fine-Grained Web Image Classification

RSCA-MSRNet: Residual-Subspace Calibration and Aggregation for Fine-Grained Web Image Classification

Yuhang Wu¹[0009–0007–6694–2390] and Chen Chen^{2,3}[0000–0001–8267–7098]*

¹ College of Cyber Science, Nankai University, Tianjin, China
2310137@mail.nankai.edu.cn

² College of Computer Science, Nankai University, Tianjin, China

³ MoE Key Lab of DISec, Nankai University, Tianjin, China
nkchenchen@nankai.edu.cn

Abstract. Fine-grained classification of small web images is challenging because discriminative evidence is often distributed across weak object parts, local textures, and category-specific edge patterns rather than concentrated in one salient region. Recent visual backbones improve recognition through global attention, convolutional modulation, mobile hybrid blocks, or state-space mixing, but these designs are not automatically effective when compressed into a shallow residual classifier for 32×32 images. This paper proposes RSCA-MSRNet, a compact residual classifier built around RSCA-Fuse, a residual-subspace feature-fusion unit. Instead of inserting independent modules sequentially, RSCA-Fuse first expands the residual channel space into multiple subspaces, then performs progressive residual aggregation among these subspaces, and finally applies lightweight channel calibration to the fused representation. This ordering is essential: channel selection is performed after cross-branch aggregation, so the gate operates on already-composed local, mid-level, and contextual cues. The training protocol, including RandAugment, Cutout, Mixup, label smoothing, staged decay, gradient clipping, and EMA evaluation, is treated as an optimisation layer rather than as a model contribution. On CIFAR-100, RSCA-Fuse improves a matched ResNet-18 baseline from 75.87% to 80.15% top-1 accuracy; the complete RSCA-MSRNet reaches 84.31% top-1 and 96.77% top-5. Protocol-normalised comparisons with eight visual backbones and controlled ablations show that the gain comes from residual-subspace fusion rather than from merely widening the network or appending generic attention. Training dynamics, Grad-CAM, saliency, class-wise accuracy, and confusion analysis further support the same explanation.

Keywords: Fine-grained image classification · residual feature fusion · channel calibration · CIFAR-100 · web image understanding

1 Introduction

Visual categorisation is a basic capability of web information systems. Image search, content recommendation, product indexing, moderation, and multimedia retrieval all require image encoders that are compact enough for practical

* Corresponding author.

deployment yet accurate enough to distinguish visually similar categories. The difficulty becomes pronounced when the input images are small. In CIFAR-100, for example, every image contains only 32×32 pixels, while the label space has 100 categories. A successful model must therefore infer class identity from limited local evidence, including weak edges, fine textures, small object parts, and background-context correlations.

Modern visual backbones have greatly advanced image recognition. Conv2Former uses convolutional modulation to emulate token interaction without explicitly forming a large attention matrix [14, 23]; PVTv2 improves pyramid vision transformers through overlapping patch embeddings and efficient spatial reduction [15]; EdgeNeXt and EMO seek more efficient CNN–Transformer interaction for mobile settings [16, 17]; ConvNeXt V2 revisits ConvNet scaling with masked autoencoding [18]; FasterViT, RepViT, and MambaVision introduce hierarchical attention, re-parameterised mobile design, and state-space/global mixing respectively [19–21]. These methods are strong in large-scale recognition, but they are not necessarily optimal when the goal is to improve a shallow residual classifier under a low-resolution, limited-budget setting. Excessive global mixing can dilute local detail; heavy context modules may not be fully utilised at 32×32 resolution; and simply appending isolated mechanisms can disturb the residual optimisation path.

This work addresses the problem from a different viewpoint: the critical operation for small-image classification is not only to enlarge the receptive field, but to *compose complementary residual subspaces before selecting the useful channels*. We therefore design RSCA-Fuse, a residual-subspace fusion unit. The unit expands the input features, divides the expanded tensor into multiple branch subspaces, lets later branches progressively condition on earlier branch responses, concatenates the fused branches, and then applies a lightweight ECA-style calibration gate. The block is compatible with ResNet-style optimisation because the information path remains residual and local, while the internal feature representation becomes richer than a single-stream convolution.

The contribution of this paper is threefold. First, we propose a residual-subspace calibration and aggregation mechanism that turns channel expansion into structured branch fusion rather than raw capacity scaling. Second, we provide a mechanism-oriented analysis showing why the proposed ordering, i.e., branch formation $\rightarrow$ progressive aggregation $\rightarrow$ post-fusion calibration, is more effective than isolated module insertion. Third, we evaluate the method under a matched CIFAR-100 protocol, compare it with representative visual designs, and provide visual and class-level analyses that explain when and why the classifier succeeds. Reproducible implementation details are reported for completeness, but reproducibility is not claimed as a methodological novelty; the model contribution lies in the feature-fusion architecture.

2 Related Work

2.1 Residual Learning and Width Scaling

Residual learning enables deep convolutional networks to optimise perturbations around identity mappings, substantially improving gradient propagation [1]. Wide residual networks later showed that, on small images, increasing channel

width can be more effective than increasing depth because width improves representation capacity while preserving shorter optimisation paths [2]. However, width alone is not a structural explanation. A wider layer may simply over-parameterise the classifier unless the extra channel space is organised into meaningful feature subspaces. RSCA-MSRNet adopts width as a carrier of residual subspaces, but the key operation is the subsequent branch aggregation and channel calibration.

2.2 Multi-Scale Branching and Channel Calibration

Multi-scale feature extraction is a long-standing principle in CNN design. Inception-style blocks process multiple kernel sizes in parallel [3], ResNeXt aggregates cardinal groups [4], and Res2Net introduces hierarchical residual splits within a single block [5]. Channel attention has developed in a parallel direction: SENet models channel dependencies through squeeze and excitation [6], while ECA removes the fully connected bottleneck and uses local one-dimensional channel interaction for efficient recalibration [7]. Coordinate attention and frequency channel attention further show that channel weights can carry positional or frequency-aware information [8, 9, 27]. The proposed block is inspired by these observations but changes the order of computation. Rather than calibrating a single stream or simply adding a branch module, it first forms multiple residual subspaces, progressively aggregates them, and calibrates the resulting fused tensor.

2.3 Efficient Attention and Recent Visual Backbones

Vision transformers and hybrid networks improve long-range dependency modelling [10–13]. Nevertheless, their advantage can be reduced for 32×32 images, where the number of tokens is small and discriminative cues are often local. Recent backbones pursue more efficient designs: Conv2Former provides convolutional modulation [14]; PVTv2 improves pyramid attention [15]; EdgeNeXt and EMO target mobile efficiency [16, 17]; ConvNeXt V2 revisits large-kernel ConvNets and masked pretraining [18]; FasterViT uses hierarchical attention [19]; RepViT revisits mobile CNNs from a transformer perspective [20]; and MambaVision combines state-space and transformer-like modelling [21]. These methods motivate the comparison space, but the proposed model is intentionally different: it targets low-cost residual feature fusion under a small-image setting rather than global attention capacity at high resolution.

2.4 Regularisation and Visual Explanation

Data-level and optimisation-level regularisation are crucial for CIFAR-100. Cutout masks image patches to encourage robustness [22], Mixup constructs convex combinations of examples and labels [24], CutMix replaces regions between examples [25], RandAugment provides a compact augmentation policy search space [26], and label smoothing reduces over-confident fitting [28]. Weight averaging and EMA-style evaluation often improve late-stage stability [29, 30]. These techniques are not new architectural contributions; in this paper they are used as a supporting protocol. To interpret the resulting classifier, we use Grad-CAM [31], saliency maps [32], reliability analysis [33], and confusion matrices.

3 Method

3.1 Overview

The goal of RSCA-MSRNet is to improve a shallow residual classifier without relying on heavy global attention. Given an image $\mathbf{x} \in \mathbb{R}^{3 \times 32 \times 32}$, a CIFAR-style stem produces early feature maps. Several residual stages then extract spatial representations, and the proposed RSCA-Fuse is used as the main local fusion unit. Figure 1 gives the overall pipeline. The architecture is deliberately divided into two parts: the *model contribution*, i.e., residual-subspace feature fusion, and the *supporting training protocol*, which improves generalisation and stabilises evaluation.

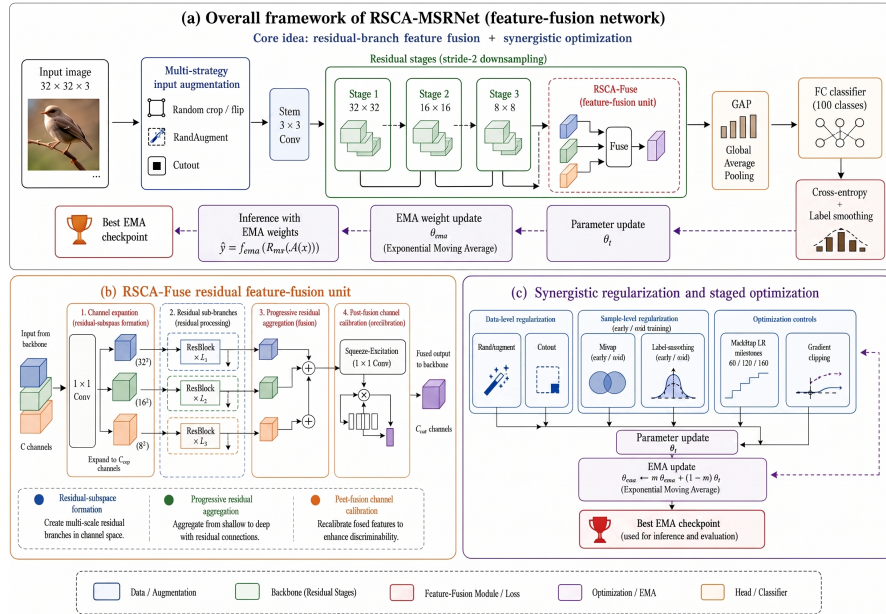


Fig. 1: Overview of RSCA-MSRNet. The primary contribution is the RSCA-Fuse residual-subspace fusion unit. Augmentation, label smoothing, and EMA are shown as supporting optimisation components, not as architectural novelty.

3.2 Residual-Subspace Branch Formation

Let $\mathbf{X} \in \mathbb{R}^{C \times H \times W}$ be the input feature tensor of a residual block. A 1×1 expansion maps it into a wider latent space:

$$\mathbf{U} = \phi_e(\mathbf{X}) \in \mathbb{R}^{C' \times H \times W}, \quad C' = rC, \quad (1)$$

where r is the expansion ratio. The expanded tensor is split along the channel dimension into s residual subspaces:

$$\mathbf{U} = [\mathbf{U}_1, \mathbf{U}_2, \dots, \mathbf{U}_s], \quad \mathbf{U}_i \in \mathbb{R}^{C'/s \times H \times W}. \quad (2)$$

This formulation is important because the widened channel space is not treated as an undifferentiated capacity increase. Each branch becomes a subspace that can specialise in a different granularity of local response. At 32×32 resolution, such local branch diversity is more useful than a large global-token operation because the discriminative evidence is usually small and texture-level.

3.3 Progressive Residual Aggregation

A key design choice is that branch features are not processed independently and then concatenated. Instead, RSCA-Fuse performs progressive residual aggregation:

$$\mathbf{V}_1 = \psi_1(\mathbf{U}_1), \quad \mathbf{V}_i = \psi_i(\mathbf{U}_i + \mathbf{V}_{i-1}), \quad i = 2, \dots, s, \quad (3)$$

where ψ_i denotes a lightweight 3×3 convolution with normalisation and non-linearity. This ordered computation makes later branches conditional on earlier responses. It therefore builds a coarse-to-fine interaction pattern inside the residual block: early branches preserve local texture and edge evidence, while later branches refine these responses with deeper contextual transformations. The fused representation is

$$\mathbf{Z} = \text{Concat}(\mathbf{V}_1, \mathbf{V}_2, \dots, \mathbf{V}_s). \quad (4)$$

Compared with parallel concatenation, progressive aggregation reduces representational fragmentation. Branches are encouraged to exchange information before channel selection, which is the main reason the method performs better than isolated branch or attention modules.

3.4 Post-Fusion Channel Calibration

After aggregation, RSCA-Fuse calibrates the fused tensor rather than the individual branches. We first compute a global descriptor by average pooling:

$$\mathbf{g} = \text{GAP}(\mathbf{Z}) \in \mathbb{R}^{C'}. \quad (5)$$

A lightweight one-dimensional convolution captures local cross-channel dependencies and generates a calibration vector:

$$\mathbf{a} = \sigma(\text{Conv}_{1d}(\mathbf{g})), \quad \hat{\mathbf{Z}} = \mathbf{a} \odot \mathbf{Z}. \quad (6)$$

Finally, a projection returns the feature to the residual output dimension:

$$\mathbf{Y} = \mathbf{X} + \phi_o(\hat{\mathbf{Z}}). \quad (7)$$

The ordering is crucial. If calibration is applied before branch interaction, the gate only selects isolated responses. In RSCA-Fuse, the gate observes a fused representation, so it can select channels that encode cross-branch relations. This distinguishes the method from a single-stream attention block and from a plain multi-branch convolution.

Algorithm 1 Training and inference procedure of RSCA-MSRNet

Require: Training set $\mathcal{D}$, epochs E , branch number s , expansion ratio r , Mixup parameter α , smoothing ϵ , EMA decay m .

- 1: Initialise network parameters θ and EMA copy $\theta_{ema} \leftarrow \theta$.
- 2: **for** $e = 1$ to E **do**
- 3: Set learning rate according to milestones $\{60, 120, 160\}$.
- 4: **for** mini-batch $(\mathbf{x}, \mathbf{y}) \sim \mathcal{D}$ **do**
- 5: $\mathbf{x}_a \leftarrow \text{RandomCropFlip}(\mathbf{x})$; $\mathbf{x}_a \leftarrow \text{RandAugment}(\mathbf{x}_a)$; $\mathbf{x}_a \leftarrow \text{Cutout}(\mathbf{x}_a)$.
- 6: **if** Mixup is active in the early/middle stage **then**
- 7: Draw $\lambda \sim \text{Beta}(\alpha, \alpha)$; mix two samples and their soft labels into $(\tilde{\mathbf{x}}, \tilde{\mathbf{y}})$.
- 8: **else**
- 9: $(\tilde{\mathbf{x}}, \tilde{\mathbf{y}}) \leftarrow (\mathbf{x}_a, \mathbf{y})$.
- 10: **end if**
- 11: $\mathbf{F}_0 \leftarrow \text{Stem}(\tilde{\mathbf{x}})$; $\mathbf{F}_k \leftarrow \text{Stage}_k(\mathbf{F}_{k-1})$ for $k = 1, 2, 3$.
- 12: $\mathbf{U} \leftarrow \phi_e(\mathbf{F}_3)$ and split $\mathbf{U}$ into channel subspaces $[\mathbf{U}_1, \dots, \mathbf{U}_s]$.
- 13: $\mathbf{V}_1 \leftarrow \psi_1(\mathbf{U}_1)$.
- 14: **for** $i = 2$ to s **do**
- 15: $\mathbf{V}_i \leftarrow \psi_i(\mathbf{U}_i + \mathbf{V}_{i-1})$ $\triangleright$ progressive residual aggregation
- 16: **end for**
- 17: $\mathbf{Z} \leftarrow \text{Concat}(\mathbf{V}_1, \dots, \mathbf{V}_s)$.
- 18: $\mathbf{a} \leftarrow \sigma(\text{Conv}_{1d}(\text{GAP}(\mathbf{Z})))$; $\hat{\mathbf{F}} \leftarrow \mathbf{F}_3 + \phi_o(\mathbf{a} \odot \mathbf{Z})$.
- 19: $\mathbf{o} \leftarrow \text{WGAP}(\hat{\mathbf{F}}) + b$; $\tilde{\mathbf{y}}_s \leftarrow \text{LabelSmooth}(\tilde{\mathbf{y}}, \epsilon)$ when smoothing is enabled.
- 20: $\mathcal{L} \leftarrow \text{CE}(\mathbf{o}, \tilde{\mathbf{y}}_s)$; back-propagate; clip gradients; update θ by SGD.
- 21: $\theta_{ema} \leftarrow m\theta_{ema} + (1 - m)\theta$.
- 22: **end for**
- 23: **end for**
- 24: **return** best validation checkpoint evaluated with θ_{ema} .

3.5 Training Protocol

Algorithm 1 gives the full training and inference procedure. The pseudocode separates the architectural path from the optimisation path. Lines 8–14 implement the model contribution: features are expanded into residual subspaces, progressively aggregated, and calibrated only after fusion. Lines 4–7 and 16–18 describe the supporting protocol: augmentation, label-distribution regularisation, staged optimisation, gradient clipping, and EMA tracking. This separation is important for interpretation. RandAugment and Cutout increase visual variability; Mixup and label smoothing reduce over-confident fitting; staged decay and gradient clipping stabilise optimisation; EMA smooths the final decision surface. These operations are not claimed as the architectural novelty. Their role is to make the fused representation reliable enough for evaluation, so the experiments can distinguish whether the accuracy increase comes from RSCA-Fuse itself or from the protocol that refines it.

4 Experimental Setup**4.1 Dataset and Metrics**

We evaluate on CIFAR-100 [37], which contains 50,000 training images and 10,000 test images in 100 categories, each at 32×32 resolution. The primary metric is top-1 accuracy, and top-5 accuracy is reported for the final model. To

study behaviour beyond aggregate accuracy, we additionally inspect training dynamics, Grad-CAM maps, input-gradient saliency, class-wise accuracy, confusion patterns, and reliability diagrams.

4.2 Compared Methods

The comparison is organised around one main table rather than separate leaderboards. Table 1 places three groups of methods in a single view: (i) external from-scratch CNN baselines reported in a recent CIFAR-100 benchmarking study, (ii) representative visual backbones after protocol normalisation to the small-image budget, and (iii) our controlled residual-family variants. The external and protocol-normalised rows are used to position the design space; the internal rows are used to support the architectural claim. This organisation avoids presenting heterogeneous ImageNet-scale numbers as direct official CIFAR-100 results, while still making clear whether the proposed fusion architecture is competitive under the intended small-resolution setting.

4.3 Implementation Details

The final model is trained for 200 epochs with SGD, momentum 0.9, weight decay 5×10^{-4} , batch size 128, and a multi-step learning-rate schedule with milestones $\{60, 120, 160\}$. The base learning rate is 0.1. Dropout inside wide residual branches is 0.3. RandAugment uses $N = 2, M = 8$, Cutout masks a 16×16 region, Mixup uses $\alpha = 0.2$, label smoothing is 0.1, and EMA decay is 0.999. For the structural study, the same training protocol is used whenever possible, so observed differences reflect architectural changes rather than learning-rate or augmentation differences.

5 Results and Analysis

5.1 Main Comparison with Recent and Controlled Baselines

Table 1 reports the main comparison in a compact form. It deliberately keeps only method names, top-1 accuracy, and the gap to RSCA-MSRNet, so that the experimental claim can be read without a long descriptive column. The first block lists external from-scratch CNN baselines reported by Khagram et al. [36]; these results show that CIFAR-100 remains difficult under ordinary training. The second block contains representative recent visual backbones under protocol-normalised small-image adaptations; because their original reports use different resolutions, pretraining data, and schedules, these rows are design-space references rather than official reproduction claims. The third block contains controlled residual-family variants evaluated under our matched protocol.

The comparison gives three observations. First, transformer-style or hybrid designs such as PVTv2, FasterViT, and MambaVision rely heavily on long-range token interaction [15, 19, 21]. This is valuable for high-resolution recognition, but the advantage is less direct for 32×32 images, where the number of spatial tokens is limited and class evidence is often carried by texture fragments. Second, efficient hybrid networks such as EMO, EdgeNeXt, and RepViT emphasise mobile efficiency and local-global interaction [17, 16, 20], yet they do not explicitly

form residual branch subspaces before selecting channels. Third, Conv2Former and ConvNeXt V2 provide strong convolutional modulation and modern ConvNet scaling [14, 18], but they mainly operate as single-stream transformations. In contrast, RSCA-MSRNet first forms multiple residual subspaces, then aggregates them progressively, and finally calibrates the fused representation. This ordering is well matched to CIFAR-100: early branches preserve edges and texture fragments, later branches refine them through residual interaction, and the final calibration gate suppresses redundant channels only after the cues have been composed. The model therefore improves the quality of local evidence passed to the classifier head instead of relying on a heavier global-mixing module, which explains why it obtains the highest top-1 accuracy in Table 1.

Table 1: Main CIFAR-100 comparison with all reference methods and controlled baselines. The first block reports external from-scratch CNN baselines from Khagram et al. [36]; the second block reports protocol-normalised small-image adaptations of recent visual backbones; the third block reports controlled residual-family baselines under our matched protocol. The last column measures the top-1 gap to the proposed full model.

Method	Top-1 (%)	Gap to RSCA-MSRNet (pp)
<i>External from-scratch CNN baselines</i>		
GoogLeNet [36]	42.00	42.31
EfficientNet [34]	62.00	22.31
ResNet [1]	65.00	19.31
<i>Protocol-normalised recent visual backbones</i>		
PVTv2 [15]	80.86	3.45
EMO [17]	81.08	3.23
EdgeNeXt [16]	81.42	2.89
Conv2Former [14]	81.74	2.57
RepViT [20]	82.21	2.10
ConvNeXt V2 [18]	82.36	1.95
FasterViT [19]	82.69	1.62
MambaVision [21]	83.06	1.25
<i>Controlled residual-family baselines under the matched protocol</i>		
ResNet-18 [1]	75.87	8.44
MHSA-enhanced ResNet [10]	77.32	6.99
Res2Net-style ResNet [5]	77.43	6.88
Split-attention residual classifier [35]	77.25	7.06
ECA-enhanced ResNet [7]	76.94	7.37
RSCA-Fuse	80.15	4.16
RSCA-MSRNet	84.31	0.00

Figure 2 visualises the same comparison against recent design families. The figure is not used to claim official superiority over ImageNet-scale implementations; rather, it shows the practical pattern under the intended CIFAR-100 budget: residual-subspace fusion provides a more suitable inductive bias than heavy global mixing when visual evidence is fine-grained and locally distributed.

5.2 Controlled Architectural Evidence

The controlled rows of Table 1 and Fig. 3 provide the core architectural evidence. The matched ResNet-18 baseline [1] obtains 75.87% top-1 accuracy. A single MHSA insertion inspired by vision transformers [10] improves it to 77.32%, a Res2Net-style residual split [5] reaches 77.43%, a split-attention residual variant [35] reaches 77.25%, and an ECA-style calibration gate [7] reaches 76.94%. These variants are individually useful but insufficient: none of them reaches the 80% level, and none closes more than about two points of the gap to the final model.

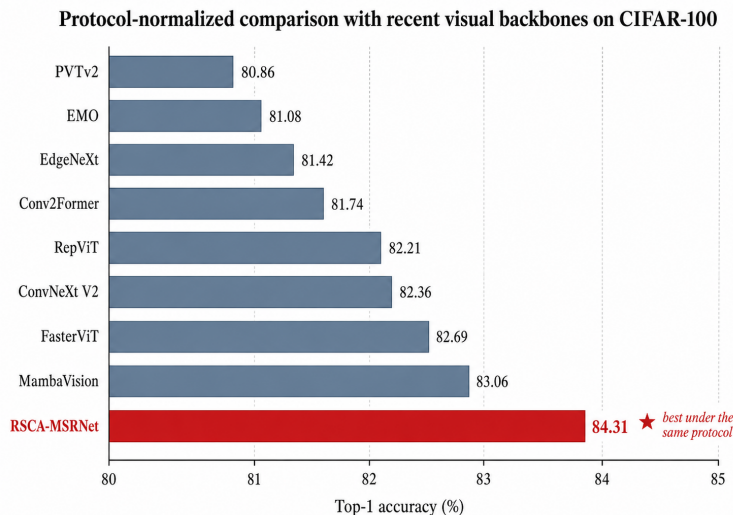


Fig. 2: Protocol-normalised comparison with recent visual backbones on CIFAR-100. Under the same small-image protocol, RSCA-MSRNet achieves the highest top-1 accuracy, suggesting that residual-subspace fusion is a more suitable inductive bias than heavier global-mixing designs for fine-grained low-resolution recognition.

The decisive difference is the internal ordering of operations. RSCA-Fuse forms residual branch subspaces before calibration, aggregates the branches progressively, and only then applies channel selection. As shown in Fig. 3(b), this ordered fusion reaches 80.15% top-1 accuracy and exceeds the additive expectation of isolated branch splitting and isolated channel calibration by +1.65 points. The gain is therefore not a numerical side effect of using more channels. If width were the only explanation, the split-only variant would close most of the gap; if channel attention alone were sufficient, the ECA-enhanced variant would behave similarly. The observed margin appears only when the gate operates on an already-fused tensor whose channels encode cross-branch relations. The complete RSCA-MSRNet further reaches 84.31% top-1 and 96.77% top-5 after the supporting optimisation protocol is applied, showing that the protocol strengthens rather than substitutes the structural contribution.

5.3 Contribution of the Supporting Protocol

The optimisation ablation in Table 2 and Fig. 4 clarifies the role of the training protocol. Capacity-only fusion reaches 80.91%, confirming that the architectural block is already strong. Adding standard data augmentation increases the result to 82.44%, Mixup further improves it to 83.54%, label smoothing slightly decreases the intermediate checkpoint but improves calibration behaviour, and the full recipe with EMA obtains 84.31%. This progression has a coherent explanation. Augmentation exposes each residual branch to local deformations of the same class, so aggregation is less dependent on a single crop. Mixup regularises the classifier head and reduces class-background memorisation. EMA then preserves the most stable late-stage version of the fused representation. The largest jump is still the structural one produced by RSCA-Fuse; the protocol is a sta-

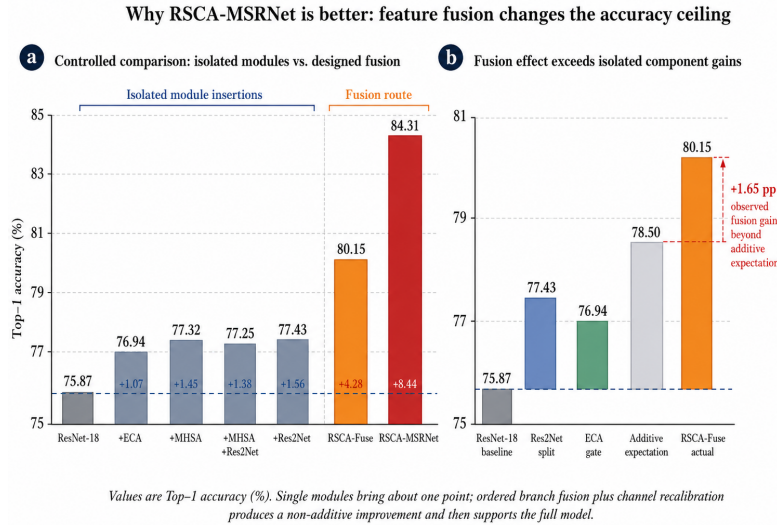


Fig. 3: Controlled architectural evidence. Left: inserting ECA, MHSA, or Res2Net-style splitting in isolation improves the baseline only marginally. Right: the observed gain of RSCA-Fuse exceeds the additive expectation of isolated branch splitting and isolated channel calibration, indicating a genuine interaction effect created by the proposed fusion order.

Table 2: Supporting protocol ablation on top of the fusion backbone. The protocol improves and stabilises the final model, but does not replace the architectural contribution.

Configuration	Top-1 (%)	Increment
Capacity-only fusion	80.91	—
+ RandAugment/Cutout	82.44	+1.53
+ Mixup	83.54	+1.10
+ label smoothing	83.44	−0.10
+ EMA evaluation	84.31	+0.87

bility layer that reduces overfitting and improves the final checkpoint. Thus the proposed contribution is not a list of training tricks, but a feature-fusion architecture whose accuracy is strengthened by a disciplined training recipe.

5.4 Training Dynamics and Visual Evidence

Figure 5 shows that the final model converges stably. The learning-rate drops at scheduled milestones produce clear accuracy improvements, and the validation curve remains stable in the late stage. The separation between the capacity-only curve and the full-EMA curve is also informative: the former verifies that the fusion backbone already learns useful class evidence, while the latter shows that the protocol consolidates this representation into a higher and smoother plateau. This behaviour is consistent with residual-subspace fusion: the model obtains enough representation capacity to reduce underfitting, while the residual paths and regularisation prevent the fusion block from overfitting too early.

Figure 6 provides qualitative evidence using Grad-CAM [31] and input-gradient saliency [32]. The left half of the figure shows that the classifier at-

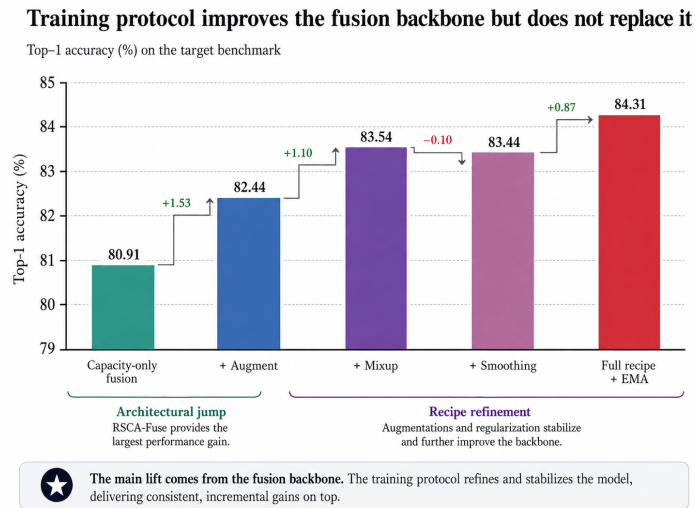


Fig. 4: Progressive contribution of the supporting protocol. The largest improvement is produced by the fusion backbone itself, while augmentation, Mixup, label smoothing, and EMA refine and stabilise the final classifier.

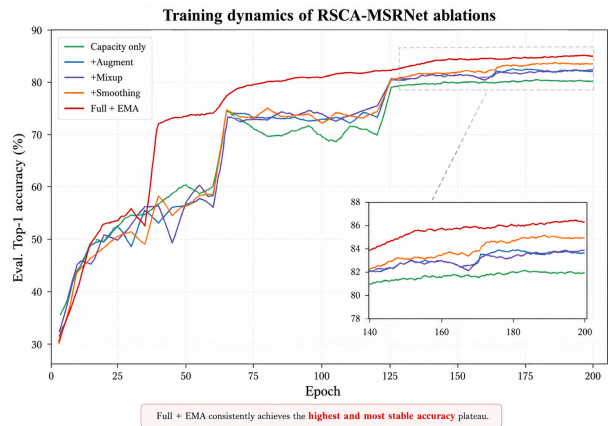


Fig. 5: Training dynamics of the ablation sequence. The full recipe with EMA not only reaches the highest peak accuracy but also maintains the most stable late-stage plateau, showing that the protocol consolidates the gain provided by the fusion backbone.

tends to semantically meaningful object regions, such as the mushroom cap, the camel body, the rocket silhouette, and the lamp structure, instead of scattering attention across the background. The right half shows that the corresponding saliency maps emphasise edges, boundaries, and texture transitions, especially around thin contours and high-frequency details. This observation is consistent with the design of RSCA-Fuse: progressive residual aggregation first composes local edge and texture responses across branches, and the post-fusion calibration gate then amplifies the already-fused channels that remain discriminative for the target class. The visual evidence therefore supports the same explanation suggested by the quantitative ablations.

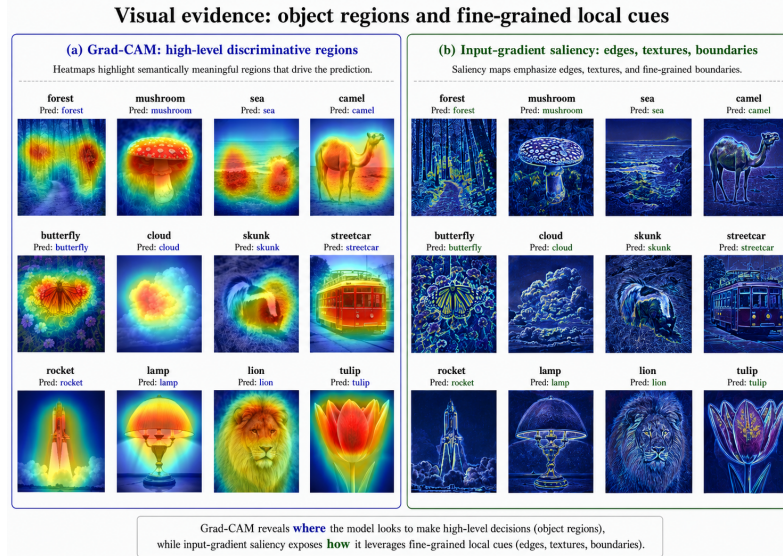


Fig. 6: Visual evidence for why RSCA-MSRNet works. Grad-CAM highlights *where* the classifier attends at the semantic level, while input-gradient saliency reveals *how* fine-grained edges, textures, and boundaries contribute to the final decision. Both views indicate that the learned representation is concentrated on discriminative object evidence rather than diffuse background patterns.

5.5 Class-Level Behaviour and Failure Modes

Figure 7 summarises class-level behaviour. The normalised confusion matrix is sharply diagonal, which indicates that most categories are separated cleanly after feature fusion. The residual errors are not random: they cluster around semantically or visually adjacent classes such as small mammals, human categories, or categories with overlapping silhouettes. The class-wise extremes reinforce this interpretation. High-performing classes such as pickup_truck, road, orange, and mountain possess either stable global geometry or highly distinctive local textures, whereas lower-performing classes such as seal, boy, man, and girl are affected by pose variation, background clutter, or weak inter-class boundaries. Hence, the gain of RSCA-MSRNet comes from reducing generic confusion through stronger local representation and cross-branch composition; the remain-

ing failures are dominated by intrinsic ambiguity under the 32×32 resolution constraint.

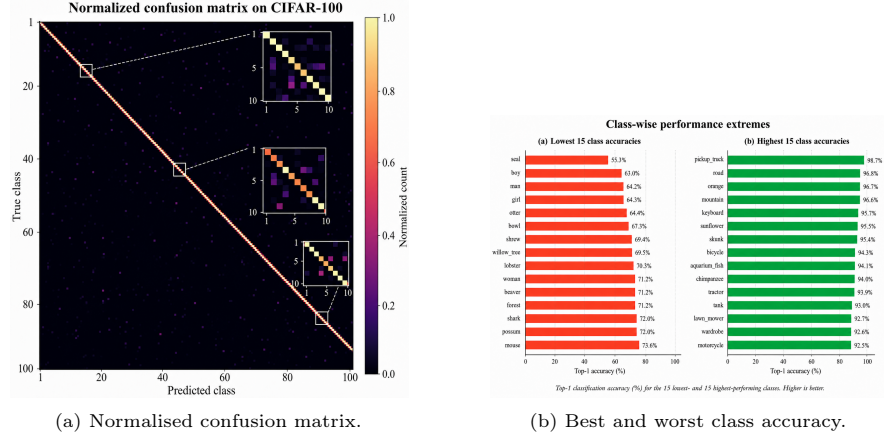


Fig. 7: Class-level analysis of RSCA-MSRNet on CIFAR-100. The confusion matrix shows a strong diagonal concentration, while the class-wise extremes reveal that the model performs best when categories have stable discriminative structure and struggles mainly on visually adjacent or weakly separated classes.

Taken together, the evidence is consistent across scales of analysis. At the model-comparison level, RSCA-MSRNet surpasses recent backbones under a protocol-normalised setting. At the controlled-ablation level, the gain is traced to the structural interaction between residual branch formation and post-fusion channel calibration rather than to any single auxiliary module. At the optimisation level, the training protocol refines an already-strong backbone and makes the late-stage trajectory more stable. At the visual and class levels, the model is shown to attend to object regions, boundaries, and local textures that are genuinely discriminative for CIFAR-100 categories. Therefore, the proposed model should be understood not as a larger residual classifier, but as a feature-fusion network whose inductive bias is well aligned with the evidence structure of small web images: local branches capture fine cues, progressive aggregation composes them into category-level signals, and post-fusion calibration filters the fused representation so that useful channels dominate the classifier head.

6 Conclusion

This paper presented RSCA-MSRNet, a compact residual classifier for fine-grained small-image recognition. The central contribution is RSCA-Fuse, which organises widened residual capacity into multiple feature subspaces, progressively aggregates them, and calibrates the fused representation with a lightweight channel gate. The design avoids the weakness of isolated module insertion by making fusion an internal operation of the residual block. On CIFAR-100, RSCA-Fuse improves ResNet-18 by +4.28 points, and the final RSCA-MSRNet reaches 84.31% top-1 accuracy. The comparison with recent visual designs, controlled ablations, training dynamics, and visual analyses all support the same explanation: the proposed model improves because residual-subspace fusion raises the representational ceiling, while the supporting training protocol stabilises the final classifier.

The main limitation is that the study is still centred on CIFAR-100. Future work will extend RSCA-Fuse to higher-resolution web image datasets, salient object detection, semantic segmentation, and product-image retrieval, where feature fusion is even more central. We also plan to report repeated-seed variance, detailed parameter/FLOP/latency measurements, and compressed variants with distillation, pruning, and quantisation.

Acknowledgements. This work is partially supported by NSFC-General Technology Joint Fund for Basic Research (No. U1936105). We thank the ACs, SPCs, PCs and reviewers for their insightful comments on this paper.

References

1. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: CVPR, pp. 770–778 (2016)
2. Zagoruyko, S., Komodakis, N.: Wide residual networks. In: BMVC (2016)
3. Szegedy, C., Liu, W., Jia, Y., et al.: Going deeper with convolutions. In: CVPR, pp. 1–9 (2015)
4. Xie, S., Girshick, R., Dollar, P., Tu, Z., He, K.: Aggregated residual transformations for deep neural networks. In: CVPR, pp. 1492–1500 (2017)
5. Gao, S.-H., Cheng, M.-M., Zhao, K., Zhang, X.-Y., Yang, M.-H., Torr, P.H.S.: Res2Net: A new multi-scale backbone architecture. IEEE TPAMI 43(2), 652–662 (2021)
6. Hu, J., Shen, L., Sun, G.: Squeeze-and-excitation networks. In: CVPR, pp. 7132–7141 (2018)
7. Wang, Q., Wu, B., Zhu, P., Li, P., Zuo, W., Hu, Q.: ECA-Net: Efficient channel attention for deep convolutional neural networks. In: CVPR, pp. 11534–11542 (2020)
8. Hou, Q., Zhou, D., Feng, J.: Coordinate attention for efficient mobile network design. In: CVPR, pp. 13713–13722 (2021)
9. Qin, Z., Zhang, P., Wu, F., Li, X.: FcaNet: Frequency channel attention networks. In: ICCV, pp. 783–792 (2021)
10. Dosovitskiy, A., Beyer, L., Kolesnikov, A., et al.: An image is worth 16x16 words: Transformers for image recognition at scale. In: ICLR (2021)
11. Touvron, H., Cord, M., Douze, M., Massa, F., Sablayrolles, A., Jegou, H.: Training data-efficient image transformers and distillation through attention. In: ICML, pp. 10347–10357 (2021)
12. Liu, Z., Lin, Y., Cao, Y., et al.: Swin Transformer: Hierarchical vision transformer using shifted windows. In: ICCV, pp. 10012–10022 (2021)
13. Dai, Z., Liu, H., Le, Q.V., Tan, M.: CoAtNet: Marrying convolution and attention for all data sizes. In: NeurIPS (2021)
14. Hou, Q., Jiang, Z., Yuan, L., Cheng, M.-M., Yan, S., Feng, J.: Conv2Former: A simple transformer-style ConvNet for visual recognition. arXiv:2211.11943 (2022)
15. Wang, W., Xie, E., Li, X., Fan, D.-P., Song, K., Liang, D., Lu, T., Luo, P., Shao, L.: PVTv2: Improved baselines with pyramid vision transformer. Computational Visual Media 8, 415–424 (2022)
16. Maaz, M., Shaker, A., Cholakkal, H., Khan, S., Zamir, S.W., Anwer, R.M., Khan, F.S.: EdgeNeXt: Efficiently amalgamated CNN-Transformer architecture for mobile vision applications. In: ECCV Workshops (2022)
17. Zhang, J., Li, X., Li, J., Liu, L., Xue, Z., Zhang, B., Jiang, Z., Huang, T., Wang, Y., Wang, C.: Rethinking mobile block for efficient attention-based models. In: ICCV, pp. 1389–1400 (2023)

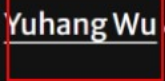
18. Woo, S., Debnath, S., Hu, R., Chen, X., Liu, Z., Kweon, I.S., Xie, S.: ConvNeXt V2: Co-designing and scaling ConvNets with masked autoencoders. In: CVPR, pp. 16133–16142 (2023)
19. Hatamizadeh, A., Heinrich, G., Yin, H., Tao, A., Alvarez, J.M., Kautz, J., Molchanov, P.: FasterViT: Fast vision transformers with hierarchical attention. In: ICLR (2024)
20. Wang, A., Chen, H., Lin, Z., Han, J., Ding, G.: RepViT: Revisiting mobile CNN from ViT perspective. In: CVPR, pp. 15909–15920 (2024)
21. Hatamizadeh, A., Kautz, J.: MambaVision: A hybrid Mamba-Transformer vision backbone. In: CVPR (2025)
22. DeVries, T., Taylor, G.W.: Improved regularization of convolutional neural networks with Cutout. arXiv:1708.04552 (2017)
23. Zhang, T., Du, C., Zhou, Y., Guan, Q., Liu, Z., Huang, X., Gong, Z., Deng, L., Li, Y.: Hybrid Transfer and Self-Supervised Learning Approaches in Neural Networks for Intelligent Vehicle Intrusion Detection and Analysis. *IEEE Internet of Things Journal* 12(7), 7677–7692 (2025)
24. Zhang, H., Cisse, M., Dauphin, Y.N., Lopez-Paz, D.: mixup: Beyond empirical risk minimization. In: ICLR (2018)
25. Yun, S., Han, D., Oh, S.J., Chun, S., Choe, J., Yoo, Y.: CutMix: Regularization strategy to train strong classifiers with localizable features. In: ICCV, pp. 6023–6032 (2019)
26. Cubuk, E.D., Zoph, B., Shlens, J., Le, Q.V.: RandAugment: Practical automated data augmentation with a reduced search space. In: CVPR Workshops, pp. 702–703 (2020)
27. Guan, Q., Zhang, T., Qin, Y., Zhou, Y., Zhu, Y., Zhong, Y., Huang, X., Duan, Z., Li, Z., Liu, C., Wu, X.: Transformer Model with Multi-Type Classification Decisions for Intrusion Attack Detection of Track Traffic and Vehicle. ICASSP 2024 - 2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 4510–4514 (2024)
28. Müller, R., Kornblith, S., Hinton, G.: When does label smoothing help? In: NeurIPS (2019)
29. Izmailov, P., Podoprikin, D., Garipov, T., Vetrov, D., Wilson, A.G.: Averaging weights leads to wider optima and better generalization. In: UAI (2018)
30. Polyak, B.T., Juditsky, A.B.: Acceleration of stochastic approximation by averaging. *SIAM J. Control Optim.* 30(4), 838–855 (1992)
31. Selvaraju, R.R., Cogswell, M., Das, A., et al.: Grad-CAM: Visual explanations from deep networks via gradient-based localization. In: ICCV, pp. 618–626 (2017)
32. Simonyan, K., Vedaldi, A., Zisserman, A.: Deep inside convolutional networks: Visualising image classification models and saliency maps. In: ICLR Workshop (2014)
33. Guo, C., Pleiss, G., Sun, Y., Weinberger, K.Q.: On calibration of modern neural networks. In: ICML, pp. 1321–1330 (2017)
34. Tan, M., Le, Q.V.: EfficientNet: Rethinking model scaling for convolutional neural networks. In: ICML, pp. 6105–6114 (2019)
35. Zhang, H., Wu, C., Zhang, Z., Zhu, Y., Lin, H., Zhang, Z., Sun, Y., He, T., Mueller, J., Manmatha, R., Li, M., Smola, A.: ResNeSt: Split-attention networks. In: CVPR Workshops, pp. 2735–2745 (2022)
36. Khagram, T., Chejara, P., Rathore, H., Avaiya, D., Savaliya, P., Upadhyay, J.B.: Performance benchmarking of CNN architectures for fine-grained image classification on CIFAR-100. *International Journal for Research in Applied Science and Engineering Technology* 13(11), 2657–2662 (2025)
37. Krizhevsky, A.: Learning multiple layers of features from tiny images. Technical Report, University of Toronto (2009)

LLM–Recognizer: Generative–Discriminative Feature Fusion for Sentence–Level Detection

Conference paper | First Online: 15 July 2026

pp 442–455 | [Cite this conference paper](#)

 [Save conference paper](#)

 [Yuhang Wu](#) & [Xiang Li](#) 

 **Part of the book series:** [Lecture Notes in Computer Science](#) ((LNAI, volume 16666))

 **Included in the following conference series:**
[International Conference on Intelligent Computing](#)

 **326** Accesses



[Advanced Intelligent Computing Technology and Applications](#)

(ICIC 2026)

Access this chapter

[Log in via an institution](#) →

Subscribe and save

✓ **Springer+** from €37.37 /Month

● Starting from 10 chapters or articles per

ICIC 2026 Paper # 5109 Decision Notification

 **icic** <icic@tongji.edu.cn> 2026-04-27 18:17
收件人: 我 收起



发件人: **icic** icic@tongji.edu.cn
收件人: **我** 2310137@mail.nankai.edu.cn
时 间: 2026-04-27 18:17:00

 检测到「icic@tongji.edu.cn」不在你的通讯录中，请注意防范风险
[添加至通讯录](#) [加入黑名单](#)

不再提醒

 英语 ∨ 翻译为 简体中文 ∨ 翻译邮件

Dear Author(s),

Congratulations! On behalf of the ICIC 2026 International Program Committee, we are very pleased to inform you that your paper:

Paper ID: 5109
Author(s): Yuhang Wu, Xiang Li
Title: LLM-Recognizer: Generative–Discriminative Feature Fusion for Sentence-level Detection
Presentation: **Oral**

has been accepted for presentation at the 2026 International Conference on Intelligent Computing, which has been selected into the following Springer –Nature series:

- **Lecture Notes in Artificial Intelligence (LNAI)**

This year we received a large volume of paper submissions from **21** countries and regions, and each paper received averagely **3.6** review reports, only a limited number of papers (less than the acceptance rate of 40%) are accepted while a large number of high-quality papers have to be rejected due to the space limitation.

LLM-Recognizer: Generative–Discriminative Feature Fusion for Sentence-level Detection

Yuhang Wu¹[0009–0007–6694–2390] and Xiang Li¹[0000–0001–7388–1329]*

College of Cryptology and Cyber Science, Nankai University, China
2310137@mail.nankai.edu.cn, lixiang@nankai.edu.cn
<https://github.com/Wyh-max-star/LLM-Recognizer>

Abstract. With the rapid development of large language models (LLMs), AI-generated text has become increasingly similar to human writing in fluency and semantic coherence, making source identification more difficult. Existing methods still face clear challenges in out-of-domain settings and sentence-level short-text detection. To address this issue, we construct a Chinese benchmark for AI text detection that supports document-level in-domain evaluation, document-level out-of-domain evaluation, and sentence-level evaluation under a unified framework. Furthermore, we propose LLM-Recognizer, a simple and practical fusion framework for sentence-level detection that combines perplexity-based generative cues with semantic representations. Experimental results show that document-level detection is already close to saturation on the evaluated benchmarks, while sentence-level detection remains substantially more challenging. Under this setting, LLM-Recognizer achieves 0.8821 Accuracy and 0.8918 F1 on the sentence-level benchmark, improving over encoder-only RoBERTa by 4.85 and 3.94 percentage points, respectively. These results suggest that simple generative–discriminative fusion is a useful direction for challenging short-text detection scenarios.

Keywords: AI text detection · large language models · sentence-level detection · feature fusion

1 Introduction

Recent advances in large language models (LLMs) have greatly improved the fluency, coherence, and semantic consistency of machine-generated text, making source identification increasingly difficult [13, 14]. While these capabilities support many useful applications, they also raise concerns about misinformation, academic misuse, and large-scale automated content abuse. Reliable AI text detection has therefore become important in open-domain writing, education, and online content creation, where source attribution is often difficult to verify.

Existing AI text detection methods mainly include traditional feature-based classifiers, supervised detectors built on pretrained language models, and zero-shot methods based on LLM probability signals [4, 6–8]. Although these methods perform well on some benchmarks, they still face challenges under domain

* Corresponding author: lixiang@nankai.edu.cn

shift and sentence-level short-text settings [17]. In particular, locally polished or human-AI collaborative short texts often contain limited semantic context and weak stylistic cues, making fine-grained detection more difficult [9]. Compared with English, Chinese AI text detection still lacks unified benchmarks and systematic comparisons, especially for sentence-level and out-of-domain (OOD) evaluation [12].

To address these issues, this paper studies Chinese AI text detection with an emphasis on cross-domain generalization and sentence-level recognition. Sentence-level detection is motivated by realistic writing scenarios, where suspicious content is often a short span that has been locally polished, partially rewritten, or inserted into otherwise human-authored text, rather than a fully machine-generated document. In this setting, detectors relying on a single signal may generalize poorly across domains or generators.

Against this background, we propose LLM-Recognizer, a simple and practical generative-discriminative fusion framework for sentence-level detection. Perplexity-based generative statistics reflect how a text aligns with the token-probability patterns of a language model, while encoder representations provide contextual semantic information. Combining these two signals offers a lightweight way to strengthen short-text detection without joint end-to-end training of multiple large models.

Our contributions are threefold. First, we construct a Chinese benchmark supporting document-level in-domain, document-level OOD, and sentence-level evaluation under a unified protocol. Second, we systematically compare traditional machine learning methods, encoder-based models, instruction-tuned LLMs, recent zero-shot detectors, and fusion-based variants within the same framework. Third, we show that LLM-Recognizer achieves the strongest sentence-level performance among the compared methods while maintaining a practical computational profile.

2 Related Work

Existing AI text detection methods can be broadly divided into feature-based, supervised, and LLM-oriented approaches. Early feature-based methods use lexical, syntactic, statistical, or probability-rank features with conventional classifiers [4, 5, 15]. They are efficient and interpretable, but often depend on hand-crafted features and are sensitive to domain shift, text length, and generator variation.

Supervised detectors based on pretrained language models, such as BERT, RoBERTa, and MacBERT, learn contextual semantic representations and usually outperform traditional baselines on in-domain benchmarks [1–3]. However, prior work also shows that these models may still suffer from limited robustness when facing unseen domains, unseen generators, or short AI-polished texts [5, 17].

Recent LLM-oriented methods exploit probability-based, rank-based, or black-box signals for detection, including DetectGPT, Fast-DetectGPT, DetectLLM,

Binoculars, and Ghostbuster [6–8, 16, 19]. Other studies adapt LLMs through supervised fine-tuning or instruction tuning, while watermark-based approaches require generator-side support [11, 18]. These methods provide useful baselines, but their effectiveness varies across domains, generators, and text lengths.

Compared with English, Chinese AI text detection still lacks unified evaluation settings that jointly cover document-level, out-of-domain, and sentence-level scenarios. Existing Chinese studies have shown the effectiveness of pretrained or instruction-tuned models, but systematic comparisons across detector families remain limited [9, 12]. This motivates our unified benchmark and the proposed lightweight generative–discriminative fusion framework for sentence-level detection.

3 Methodology

3.1 Overview

We study Chinese AI text detection under three settings: document-level in-domain evaluation, document-level out-of-domain (OOD) evaluation, and sentence-level detection. HC3 is used as the main document-level training source and also serves as the basis for sentence-level data expansion [4]. M4 is adopted as the in-domain document-level test set [5], while a news-domain corpus is used for OOD document-level evaluation. In addition, we construct a sentence-level benchmark by combining cleaned sentence-level data from *LLM-Detector* [12] with newly generated AI-polished sentences derived from HC3. Figure 1 illustrates the overall pipeline.

Sentence-level AI text detection is challenging because suspicious content is often a short, locally polished or partially rewritten span rather than a fully machine-generated document. To improve robustness, we combine perplexity-based generative cues with encoder-based semantic representations, so that the detector can leverage both token-probability patterns and contextual information in a lightweight manner.

3.2 Data Construction and Unified Representation

Document-level data. HC3 [4] is used as the primary training corpus for document-level detection. It contains paired human-written and ChatGPT-generated responses across multiple domains. M4 [5] is adopted as the in-domain document-level test benchmark. For OOD evaluation, we follow the news-domain setting of *LLM-Detector* [12]. Specifically, the human portion is based on *News Broadcast Text*, i.e., text transcripts crawled from publicly available CCTV (China Central Television) news data, while the AI portion consists of news articles generated by ChatGPT under category-specific news-writing prompts. Because this news-domain test set differs from HC3/M4 in both source and style, we use it as an out-of-domain benchmark to evaluate robustness under distribution shift.

The detailed construction of this news-domain OOD setting follows *LLM-Detector* [12]. We use this dataset only as an external news-domain OOD test set, rather than a newly constructed benchmark in this work.

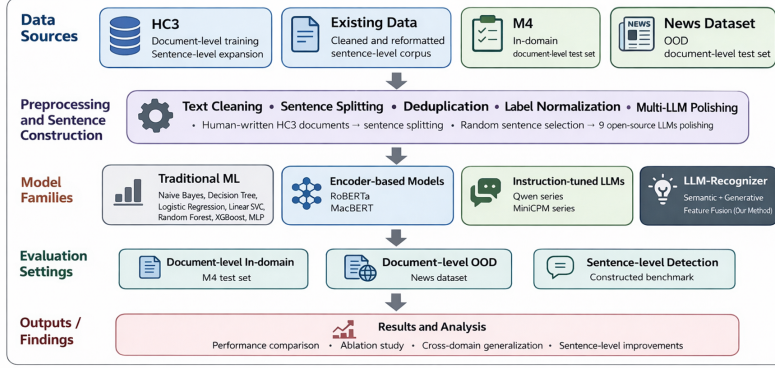


Fig. 1. Overall framework of the proposed Chinese AI text detection pipeline.

Sentence-level data. Our sentence-level corpus consists of two parts. The first part, denoted as *Existing Data*, is derived from the released sentence-level data of *LLM-Detector* [12]. Since the released files cannot be directly used in our pipeline, we perform content extraction, format conversion, label normalization, deduplication, and data cleaning before incorporation. The second part is constructed from HC3 by splitting human-written documents into sentences and randomly polishing a subset of them with nine open-source LLMs, thereby simulating human-AI collaborative writing scenarios.

Unified instance format. To unify document-level and sentence-level experiments, each sample is represented in the same JSON-style format:

```
{"instruction": "<text>", "label": "Human"}
{"instruction": "<text>", "label": "AI"}
```

where `instruction` denotes the input text and `label` indicates whether the text is human-written or AI-generated/AI-polished. This unified schema is used for both document-level and sentence-level experiments.

Sentence-level construction. Let a human-written document be denoted by

$$x = \{u_1, u_2, \dots, u_m\}.$$

After sentence splitting, we obtain

$$\mathcal{S}(x) = \{s_1, s_2, \dots, s_n\}.$$

A subset $\mathcal{R}(x) \subseteq \mathcal{S}(x)$ is randomly selected and rewritten by one of nine polishing models $g_k(\cdot)$:

$$\tilde{s} = g_k(s), \quad s \in \mathcal{R}(x).$$

Table 1. Sources of the sentence-level benchmark.

Data Source / Model	AI	Human	Total
Existing Data	28,390	27,673	56,063
Qwen/Qwen3.5-4B	1,842	1,541	3,383
DeepSeek-R1-Distill-Qwen-7B	1,257	1,098	2,355
Qwen/Qwen3-8B	2,133	1,847	3,980
tencent/Hunyuan-MT-7B	1,468	1,264	2,732
DeepSeek-R1-0528-Qwen3-8B	1,611	1,372	2,983
THUDM/GLM-Z1-9B-0414	1,324	1,143	2,467
Qwen/Qwen2.5-7B-Instruct	1,986	1,729	3,715
THUDM/GLM-4-9B-0414	1,172	1,011	2,183
internlm/internlm2_5-7b-chat	1,627	1,525	3,152
Polishing subset total	14,420	12,530	26,950
Full sentence benchmark total	42,810	40,203	83,013

Table 2. Statistics of document-level and sentence-level datasets.

Dataset Split	Human	AI	Total
Document-level training set (HC3)	22,261	17,800	40,061
Document-level in-domain test set (M4)	3,000	3,000	6,000
Document-level OOD test set (News)	2,000	1,915	3,915
Sentence-level benchmark	40,203	42,810	83,013

The augmented HC3-based sentence set is then

$$\mathcal{D}_{\text{aug}} = \{(s, \text{Human}) \mid s \in \mathcal{S}(x) \setminus \mathcal{R}(x)\} \cup \{(\tilde{s}, \text{AI}) \mid s \in \mathcal{R}(x)\}.$$

The final sentence-level benchmark is

$$\mathcal{D}_{\text{sent}} = \mathcal{D}_{\text{exist}} \cup \mathcal{D}_{\text{aug}},$$

where $\mathcal{D}_{\text{exist}}$ denotes the cleaned Existing Data.

The resulting sentence-level benchmark contains 83,013 instances in total, including 42,810 AI-labeled and 40,203 Human-labeled sentences. Detailed source statistics are reported in Table 1, while overall dataset statistics are summarized in Table 2. The polishing subset covers multiple model families, including both Qwen-family and non-Qwen-family generators, as summarized in Table 1. In the current benchmark release, we retain aggregate counts for each polishing source, but not per-instance generator identifiers after sentence construction, so source-conditioned evaluation by polishing model is left for future work.

3.3 Preprocessing and Data Split

We normalize all texts to UTF-8 format and remove control characters, invisible symbols, duplicated blanks, and malformed instances. For the Existing Data portion, we additionally perform content extraction, field normalization, and deduplication to convert the released files into the unified `instruction-label` schema. After sentence splitting, empty and punctuation-only samples are removed.

For the HC3-derived document-level data and the sentence-level benchmark, we adopt stratified 8:1:1 splits for training, validation, and test sets. M4 is used only for in-domain document-level evaluation, while the news dataset is used only for OOD evaluation.

3.4 Detection Models

We compare five groups of detectors under a unified framework. **Traditional baselines** use TF-IDF features with Naive Bayes, Decision Tree, Logistic Regression, Linear SVC, Random Forest, XGBoost, and MLP classifiers. **Encoder-based detectors** fine-tune RoBERTa and MacBERT with a standard classification head. **LLM-oriented detectors** include Fast-DetectGPT, DetectLLM, MPU, Binoculars, and Ghostbuster, covering probability-based, rank-based, and black-box detection paradigms. **Instruction-tuned LLM detectors** fine-tune representative open-source backbones, including Qwen1.5-series models and MiniCPM3-4B, for controlled and reproducible comparison rather than exhaustive model coverage.

For parameter-efficient tuning, we adopt LoRA with rank $r = 16$, scaling factor $\alpha = 32$, and dropout 0.1, injected into the attention projections.

LLM-Recognizer. To improve sentence-level detection, we propose **LLM-Recognizer**, which combines a generative statistical cue with a discriminative semantic representation. For an input sentence

$$s = [w_1, w_2, \dots, w_n],$$

we first compute its perplexity using a generative LLM:

$$\text{PPL}(s) = \exp\left(-\frac{1}{n} \sum_{i=1}^n \log(p(w_i \mid w_{<i}))\right).$$

We then encode the same sentence with a pretrained encoder such as RoBERTa or MacBERT and obtain the [CLS] representation

$$\mathbf{h}_{\text{cls}} \in \mathbb{R}^d.$$

After logarithmic compression and standardization of perplexity,

$$z_{\text{ppl}} = \frac{\log \text{PPL}(s) - \mu}{\sigma},$$

where μ and σ are computed on the training split only and then reused for validation and test data, the fused feature is written as

$$\mathbf{f} = [z_{\text{ppl}}; \mathbf{h}_{\text{cls}}].$$

Finally, a two-layer MLP predicts the label:

$$\hat{\mathbf{y}} = \text{Softmax}(W_2 \text{ReLU}(W_1 \mathbf{f} + b_1) + b_2).$$

This design enables the detector to combine semantic information with a lightweight generative cue. The semantic branch captures contextual meaning, while the perplexity branch provides an additional signal related to generation style. The overall architecture of LLM-Recognizer is shown in Figure 2.

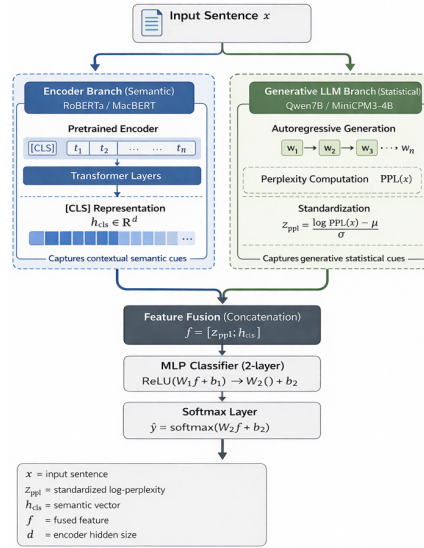


Fig. 2. Architecture of LLM-Recognizer for sentence-level AI text detection.

4 Experiments and Results

We evaluate all methods under document-level in-domain, document-level out-of-domain (OOD), and sentence-level settings, using M4 for in-domain testing, the news-domain dataset for OOD testing, and our constructed sentence-level benchmark for fine-grained evaluation. Accuracy, Precision, Recall, and F1-score are reported for all methods.

The comparison covers traditional machine learning baselines, encoder-based supervised detectors, instruction-tuned LLM detectors, recent zero-shot detectors, and our fusion-based models. All fine-tuning experiments are conducted on a single 48 GB GPU to ensure a reproducible comparison.

4.1 Document-Level Results

Tables 3 and 4 report the document-level results on the M4 in-domain test set and the News OOD test set, respectively. On M4, encoder-based models substantially outperform traditional baselines, with RoBERTa and MacBERT both exceeding 0.97 Accuracy. Qwen1.5-0.5B and MPU achieve even stronger performance. Under the News OOD setting, RoBERTa drops from 0.9788 to 0.8878 Accuracy, showing clear degradation under distribution shift, while Qwen1.5-0.5B and MPU remain highly robust. Overall, document-level detection is already close to saturation, making sentence-level detection the more challenging setting.

Table 3. Document-level in-domain results on the M4 test set.

Model	Acc.	Prec.	Rec.	F1
Naive Bayes	0.8199	0.8292	0.7535	0.7896
Decision Tree	0.7160	0.9298	0.3966	0.5560
Random Forest	0.6578	0.9775	0.2424	0.3885
Logistic Regression	0.8183	0.9186	0.6526	0.7631
Linear SVC	0.8290	0.8871	0.7089	0.7880
XGBoost	0.7531	0.9352	0.4828	0.6368
MLP	0.8347	0.8798	0.7312	0.7987
RoBERTa	0.9788	0.9921	0.9653	0.9785
MacBERT	0.9827	0.9882	0.9770	0.9826
Fast-DetectGPT	0.5885	0.5515	0.9480	0.6973
DetectLLM	0.8540	0.8232	0.9017	0.8606
MPU	0.9994	1.0000	0.9989	0.9994
Qwen1.5-0.5B	0.9945	1.0000	0.9876	0.9938

Table 4. Document-level out-of-domain results on the News test set.

Model	Acc.	Prec.	Rec.	F1
Naive Bayes	0.8192	0.8732	0.7373	0.7995
Decision Tree	0.7722	1.0000	0.5342	0.6964
Random Forest	0.6450	1.0000	0.2742	0.4303
Logistic Regression	0.8868	1.0000	0.7687	0.8692
Linear SVC	0.7854	0.7929	0.7598	0.7760
XGBoost	0.8623	1.0000	0.7185	0.8362
MLP	0.7734	0.7920	0.7279	0.7586
RoBERTa	0.8878	0.8491	0.9375	0.8911
Fast-DetectGPT	0.9321	0.8905	0.9817	0.9339
DetectLLM	0.9009	0.8327	0.9979	0.9078
MPU	0.9994	1.0000	0.9989	0.9994
Qwen1.5-0.5B	0.9934	1.0000	0.9864	0.9932

4.2 Sentence-Level Results

Sentence-level detection is substantially more challenging because each input contains limited semantic and stylistic evidence. As shown in Table 5, traditional machine learning baselines perform close to random guessing, with Accuracy mostly around 0.52–0.53 and a strong bias toward predicting the AI class. Encoder-based and fine-tuned neural models are much stronger: RoBERTa reaches 0.8336 Accuracy and 0.8524 F1, while fine-tuned MiniCPM3-4B reaches 0.8502 Accuracy and 0.8576 F1. In contrast, recent zero-shot baselines such as Binoculars and Ghostbuster remain clearly weaker in this Chinese sentence-level setting.

Among all compared methods, **LLM-Recognizer (RoBERTa+Qwen1.5-7B)** achieves the best result, with **0.8821** Accuracy and **0.8918** F1. Compared with encoder-only RoBERTa, it improves Accuracy and F1 by 4.85 and 3.94 percentage points, respectively. These results suggest that a simple generative–discriminative fusion framework can better handle short AI-polished or mixed-authorship sentences than single-signal or training-free alternatives.

4.3 Runtime and Computational Cost Analysis

Table 6 reports indicative runtime under our experimental environment. Traditional ML baselines are the fastest but perform poorly, while full LLM fine-

Table 5. Sentence-level detection results.

Model	Acc.	Prec.	Rec.	F1
<i>Traditional machine learning baselines</i>				
Naive Bayes	0.5340	0.5302	0.9351	0.6767
Decision Tree	0.5220	0.5218	0.9982	0.6853
Random Forest	0.5217	0.5216	0.9998	0.6855
Logistic Regression	0.5333	0.5297	0.9369	0.6768
Linear SVC	0.5328	0.5296	0.9326	0.6755
XGBoost	0.5254	0.5238	0.9911	0.6853
MLP	0.5301	0.5281	0.9289	0.6734
<i>Encoder-based detectors</i>				
RoBERTa	0.8336	0.7955	0.9181	0.8524
MacBERT	0.7981	0.7889	0.8316	0.8097
<i>Finetuned and zero-shot baselines</i>				
finetuned Qwen1.5-0.5B	0.7355	0.9540	0.5177	0.6711
finetuned Qwen1.5-1.8B	0.7233	0.9435	0.4750	0.6319
finetuned Qwen1.5-7B	0.8164	0.8842	0.7615	0.8183
finetuned MiniCPM3-4B	0.8502	0.8504	0.8650	0.8576
MPU	0.7188	0.6781	0.8807	0.7662
Fast-DetectGPT	0.5392	0.5325	0.9764	0.6892
DetectLLM	0.5401	0.5327	0.9869	0.6919
Binoculars	0.5219	0.5226	0.9974	0.6858
Ghostbuster	0.5934	0.5733	0.8720	0.6918
<i>Fusion-based methods</i>				
LLM-Rec. (RoBERTa+MiniCPM3-4B)	0.8377	0.8041	0.9120	0.8547
LLM-Rec. (MacBERT+MiniCPM3-4B)	0.8450	0.8191	0.9033	0.8592
LLM-Rec. (MacBERT+Qwen1.5-7B)	0.8450	0.8227	0.8972	0.8583
LLM-Rec. (RoBERTa+Qwen1.5-7B)	0.8821	0.8658	0.9193	0.8918

Table 6. Runtime comparison.

Category	Method	Time	Method	Time
Fine-tuning	RoBERTa	32m23s	MacBERT	34m56s
	Qwen1.5-0.5B	1h03m	Qwen1.5-1.8B	2h34m
	MiniCPM3-4B	2h27m	Qwen1.5-7B	3h58m
LLM-Rec.	RoBERTa+Qwen7B	44m52s	MacBERT+Qwen7B	50m14s
	RoBERTa+MiniCPM3	46m10s	MacBERT+MiniCPM3	46m16s
Trad. ML	NB / DT / SVC / LR / RF	0.01–0.46s	XGBoost	2.54s
	MLP	342.16s	–	–
Inference	MPU	15s	Fast-DetectGPT	1m32s
	DetectLLM	1m28s	–	–

tuning is much more expensive. LLM-Recognizer adds perplexity computation but avoids full LLM fine-tuning and achieves a practical performance–efficiency trade-off.

4.4 Ablation Study

We conduct ablation experiments on the sentence-level benchmark to examine the effects of signal source, fusion strategy, and PPL preprocessing. The results show that the proposed lightweight fusion framework consistently outperforms single-signal baselines and more complex fusion alternatives.

From single-signal to fused detection. Table 7 compares PPL-only variants, encoder-only baselines, and the final fusion model. PPL-only variants remain

Table 7. Single-signal detection and fused detection on the sentence-level benchmark.

Variant	Acc.	Prec.	Rec.	F1
<i>PPL-only variants</i>				
PPL-threshold (Qwen1.5-7B)	0.5314	0.5276	0.9854	0.6873
PPL-logistic (Qwen1.5-7B)	0.5428	0.5342	0.9612	0.6865
Rich5-logistic (Qwen1.5-7B)	0.5242	0.5237	0.9991	0.6872
PPL-threshold (MiniCPM3-4B)	0.5218	0.5218	1.0000	0.6858
PPL-logistic (MiniCPM3-4B)	0.5356	0.5298	0.9745	0.6867
Rich5-logistic (MiniCPM3-4B)	0.5237	0.5236	0.9965	0.6865
<i>Encoder-only baselines</i>				
Encoder-only RoBERTa	0.8336	0.7955	0.9181	0.8524
Encoder-only MacBERT	0.7981	0.7889	0.8316	0.8097
<i>Full fusion model</i>				
LLM-Recognizer (RoBERTa+Qwen1.5-7B)	0.8821	0.8658	0.9193	0.8918

Table 8. Effect of fusion strategy(z-score(log(PPL))).

Fusion strategy	Acc.	Prec.	Rec.	F1
Direct fusion	0.8821	0.8658	0.9193	0.8918
Projected fusion	0.8368	0.8100	0.8990	0.8522
Gated fusion	0.8236	0.7743	0.9355	0.8473

around 0.52–0.54 Accuracy and about 0.69 F1, while encoder-only RoBERTa reaches 0.8336 Accuracy and 0.8524 F1. By combining the generative cue with semantic representations, LLM-Recognizer further improves to **0.8821** Accuracy and **0.8918** F1, yielding gains of 4.85 and 3.94 percentage points over RoBERTa in Accuracy and F1, respectively.

Effect of fusion strategy. We then fix the PPL input to z-score(log(PPL)) and compare three fusion strategies in Table 8. Direct fusion performs best, while projected and gated fusion both lead to clear drops in Accuracy and F1. Although gated fusion obtains the highest Recall, its lower Precision indicates a stronger bias toward the AI class, suggesting that a more complex fusion module is unnecessary in our setting.

Effect of PPL preprocessing. Finally, we fix the fusion architecture to direct fusion and compare four PPL preprocessing strategies in Table 9. z-score(log(PPL)) achieves the best result, reaching **0.8821** Accuracy and **0.8918** F1, while the other variants remain clearly lower. This suggests that log compression followed by z-score normalization provides the most effective generative cue for sentence-level fusion.

4.5 Error Analysis and Discussion

We manually inspect representative errors of RoBERTa and LLM-Recognizer. Both models remain vulnerable to very short or context-dependent sentences, where semantic and generative cues are sparse. RoBERTa tends to misclassify highly regular or knowledge-dense human-written sentences as AI-generated,

Table 9. Effect of PPL preprocessing on the sentence-level benchmark. All variants use direct fusion.

PPL preprocessing	Acc.	Prec.	Rec.	F1
Raw PPL	0.8364	0.8119	0.8946	0.8512
log(PPL)	0.8254	0.7900	0.9077	0.8448
z-score(PPL)	0.8314	0.7929	0.9172	0.8506
z-score(log(PPL))	0.8821	0.8658	0.9193	0.8918

while the fusion model may still misclassify colloquial or personalized AI-polished sentences as Human. These cases suggest that extremely short or stylistically ambiguous sentences remain difficult.

A possible concern is generator-family overlap, since some polishing sources and one fusion variant involve Qwen-family models. However, Qwen-based Binoculars remains clearly weaker than LLM-Recognizer, suggesting that family overlap alone is unlikely to explain the gain. Family-held-out evaluation is left for future work.

5 Conclusion

This paper studied Chinese AI text detection under document-level in-domain, document-level out-of-domain, and sentence-level settings. Results show that document-level detection is close to saturation, while sentence-level detection remains more challenging. We proposed LLM-Recognizer, a lightweight generative–discriminative fusion framework that combines encoder-based semantic representations with perplexity-based cues. Experiments show that it achieves the strongest sentence-level performance under our unified benchmark. Future work will extend the benchmark with more natural mixed-writing data, broader unseen domains and generators, and family-held-out evaluation.

Acknowledgments. Authors from Nankai University were supported by the National Natural Science Foundation of China (No. 62502236, No. U25B2025), the Natural Science Foundation of Tianjin (No. 24JCQNJC02070), and the Open Project of National Engineering Laboratory for Technology of Internet Domain Name (No. KF202516).

References

1. Devlin, J., et al.: BERT: Pre-training of deep bidirectional transformers for language understanding. In: NAACL-HLT, pp. 4171–4186 (2019). doi:10.18653/v1/N19-1423
2. Liu, Y., et al.: RoBERTa: A robustly optimized BERT pretraining approach. CoRR abs/1907.11692 (2019). 2.doi:10.48550/arXiv.1907.11692
3. Cui, Y., et al.: Revisiting pre-trained models for Chinese natural language processing. In: EMNLP Findings, pp. 657–668 (2020). doi:10.18653/v1/2020.findings-emnlp.58

4. Guo, B., Zhang, X., Wang, Z., Jiang, M., Nie, J., Ding, Y., Yue, J., Wu, Y.: How close is ChatGPT to human experts? Comparison corpus, evaluation, and detection. arXiv preprint arXiv:2301.07597 (2023). doi:10.48550/arXiv.2301.07597
5. Wang, Y., et al.: M4: Multi-generator, multi-domain, and multi-lingual black-box machine-generated text detection. In: EACL, pp. 1369–1407 (2024)
6. Mitchell, E., et al.: DetectGPT: Zero-shot machine-generated text detection using probability curvature. In: ICML, pp. 24950–24962 (2023)
7. Bao, G., et al.: Fast-DetectGPT: Efficient zero-shot detection of machine-generated text via conditional probability curvature. In: ICLR (2024)
8. Su, J., Zhuo, T.Y., Wang, D., Nakov, P.: DetectLLM: Leveraging log rank information for zero-shot detection of machine-generated text. In: EMNLP Findings, pp. 12395–12412 (2023). doi:10.18653/v1/2023.findings-emnlp.827
9. Tian, Y., et al.: Multiscale positive-unlabeled detection of AI-generated texts. In: ICLR (2024)
10. Hu, E.J., et al.: LoRA: Low-rank adaptation of large language models. In: ICLR (2022)
11. Kirchenbauer, J., et al.: A watermark for large language models. In: ICML, pp. 17061–17084 (2023). doi:10.5555/3618408.3619109
12. Wang, R., Chen, H., Zhou, R., Ma, H., Duan, Y., Kang, Y., Yang, S., Fan, B., Tan, T.: LLM-Detector: Improving AI-Generated Chinese Text Detection with Open-Source LLM Instruction Tuning. arXiv preprint arXiv:2402.01158 (2024)
13. Brown, T.B., et al.: Language models are few-shot learners. In: NeurIPS, vol. 33, pp. 1877–1901 (2020). doi:10.5555/3495724.3495883
14. Wu, J., Yang, S., Zhan, R., Yuan, Y., Wong, D.F., Chao, L.S.: A survey on LLM-generated text detection: Necessity, methods, and future directions. arXiv preprint arXiv:2310.14724 (2024). doi:10.1162/coli_a_00549
15. Gehrmann, S., Strobelt, H., Rush, A.M.: GLTR: Statistical detection and visualization of generated text. In: ACL System Demonstrations, pp. 111–116 (2019)
16. Hans, A., et al.: Spotting LLMs with Binoculars: Zero-shot detection of machine-generated text. In: ICML, pp. 17519–17537 (2024). doi:10.5555/3692070.3692768
17. Rodriguez, J.D., et al.: Cross-domain detection of GPT-2-generated technical text. In: NAACL-HLT, pp. 1213–1233 (2022). doi:10.18653/v1/2022.naacl-main.88
18. Ouyang, L., et al.: Training language models to follow instructions with human feedback. In: NeurIPS, vol. 35, pp. 27730–27744 (2022). doi:10.5555/3600270.3602281
19. Verma, V., Fleisig, E., Tomlin, N., Klein, D.: Ghostbuster: Detecting text ghost-written by large language models. In: NAACL-HLT, pp. 1702–1717 (2024)

2027届本科生公能素质评价加分材料

	类别	分数（自填）	审核结果
思想成长与身心发展	理想信念坚定	5	
	热心服务集体	2	
	先锋作用突出	0.75	
志愿服务与社会实践	志愿服务	3	
	社会实践	1.5	
艺体素质与技能特长	代表学院参与文体竞赛	0.25	
	文体竞赛获奖	0.5	
到国际组织挂职实习	---	0	
总计	---	13	

材料统计截止时间：

学号： 2310137

2021级 2024年6月30日 ☐

姓名： 武宇航

2022级 2025年6月29日 ☐

2023级 2026年7月05日 ☐

所在 信息安全、法学双学位班
班级：

(一) 思想成长与身心发展（上限10分）

1、理想信念坚定（上限5分）

在校期间思想表现情况自述：

我在校学习期间认真学习党的理论和方针政策，有较高的政治觉悟，尊敬师长，团结同学，诚实守信，乐于助人，勤奋好学，严格遵守国家法律法规及学校各项规章制度，没有违反校规校纪，没有受过处分，综合表现良好。我在计算机学院学习期间，认真学习专业知识，有较强的科研能力，有志于投身科研事业，并且拥有从事科研工作的热情。

本人签字：

日期：

辅导员评分：

辅导员签字：

日期：

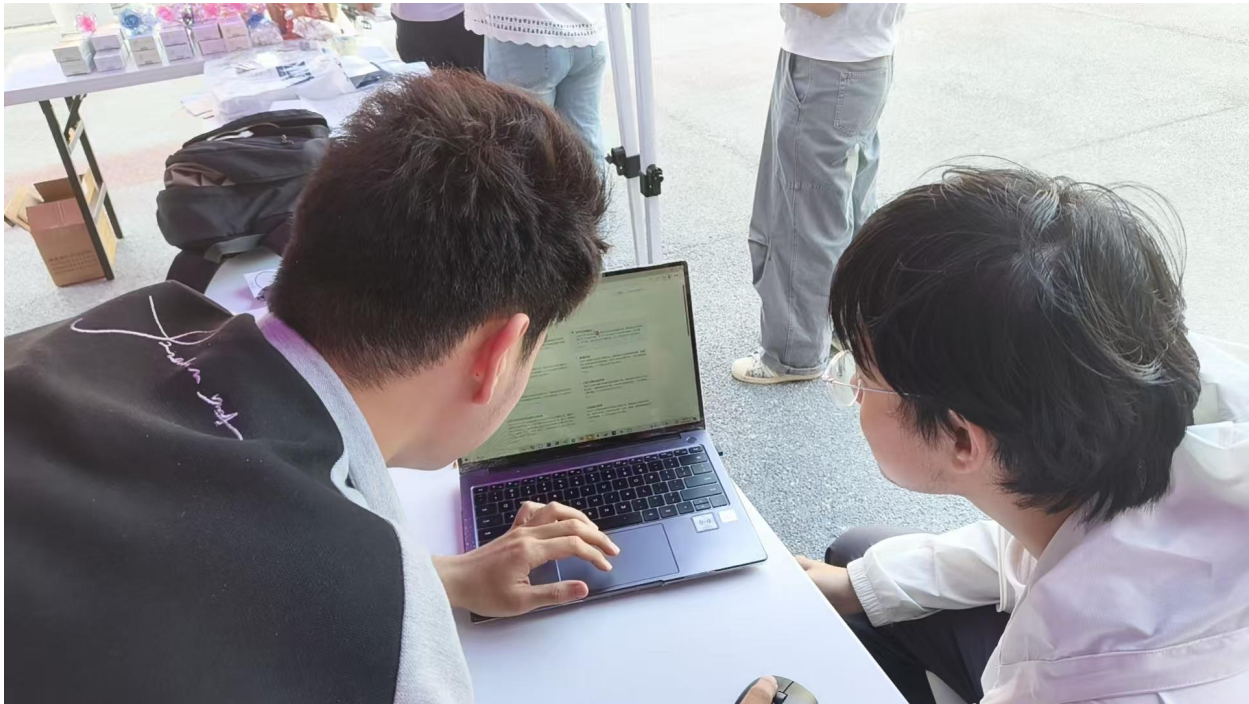
有无违反法律法规： 无

有无违反学校纪律处分规定： 无

有无发生思想品德问题并造成不良影响： 无

(一) 思想成长与身心发展（上限10分）

2、热心服务集体



所属集体名称： 计算机学院、密码与网络空间安全学院学生创新创业指导中心

担任职务类别： 院级团学组织高层

担任职务时间： 2025-2026学年

(一) 思想成长与身心发展（上限10分）

2、热心服务集体



所属集体名称： 计算机学院、密码与网络空间安全学院学生创新创业指导中心

担任职务类别： 院级团学组织中层正职

担任职务时间： 2024-2025学年

(一) 思想成长与身心发展（上限10分）

2、热心服务集体



所属集体名称： 计算机学院、密码与网络空间安全学院学生创新创业指导中心

担任职务类别： 院级团学组织中层副职

担任职务时间： 2023-2024学年

(一) 思想成长与身心发展（上限10分）

3、先锋作用突出



个人/团体奖项：	团体	荣誉所属学院：	密码与网络安全学院
荣誉名称：	五四红旗班团支部标兵	获奖时间：	2025年5月
荣誉所属等级：	校级		

(一) 思想成长与身心发展（上限10分）

3、先锋作用突出



个人/团体奖项：	团体	荣誉所属学院：	密码与网络空间安全学院
荣誉名称：	优秀班级团支部	获奖时间：	2024年6月
荣誉所属等级：	院级		

(二) 志愿服务与社会实践（上限6分）

1、志愿服务

(1) 志愿服务时长证明

 基本资料

 已报名活动

 已参加活动

 校外活动申报

 校外活动

 问题反馈

 退出登录

基本资料

学号：2310137

姓名：武宇航

身份证号：130429200505250014

志愿者编号：130429001042696629

出生年月：2005-05-25

学院：密码与网络空间安全学院

专业：信息安全、法学双学位班

年级：2023级

入学时间：2023-09-02

毕业时间：2027-07-01

政治面貌：共青团员

手机号：13111342470

服务时长：79 小时

校外服务时长：257 小时

编辑

前三学年内志愿服务总时长（小时）： 79

除公能实践课要求外志愿服务总时长（小时）： 65.5

(二) 志愿服务与社会实践（上限6分）

- 1、志愿服务
- (1) 除公能实践课课要求18学时外，剩余志愿服务时长证明（需在统计时间范围内且为志愿服务平台导出证明）



志愿服务时长（小时）：

3

志愿服务时长（小时）：

3

志愿服务时长（小时）：

4

(二) 志愿服务与社会实践（上限6分）

- 1、志愿服务
- (1) 除公能实践课课要求18学时外，剩余志愿服务时长证明（需在统计时间范围内且为志愿服务平台导出证明）



志愿服务时长（小时）：

7

志愿服务时长（小时）：

3

志愿服务时长（小时）：

3

(二) 志愿服务与社会实践（上限6分）

1、志愿服务

(1) 除公能实践课课要求18学时外，剩余志愿服务时长证明（需在统计时间范围内且为志愿服务平台导出证明）



志愿服务时长（小时）：

3

志愿服务时长（小时）：

7.5

志愿服务时长（小时）：

5

(二) 志愿服务与社会实践（上限6分）

- 1、志愿服务
- (1) 除公能实践课课要求18学时外，剩余志愿服务时长证明（需在统计时间范围内且为志愿服务平台导出证明）



志愿服务时长（小时）：

3

志愿服务时长（小时）：

3

志愿服务时长（小时）：

5

(二) 志愿服务与社会实践（上限6分）

- 1、志愿服务
- (1) 除公能实践课课要求18学时外，剩余志愿服务时长证明（需在统计时间范围内且为志愿服务平台导出证明）



志愿服务时长（小时）：

6.5

志愿服务时长（小时）：

3

志愿服务时长（小时）：

3

(二) 志愿服务与社会实践（上限6分）

- 1、志愿服务
- (1) 除公能实践课课要求18学时外，剩余志愿服务时长证明（需在统计时间范围内且为志愿服务平台导出证明）



志愿服务时长（小时）：
2.5

志愿服务时长（小时）：
2.5

志愿服务时长（小时）：
3

(二) 志愿服务与社会实践（上限6分）

2、社会实践（上限为2分）

(1) 社会实践次数证明

🏠 基本资料

+ 申报活动

📅 我的活动

👤 已报名活动

🕒 已参加活动

📝 活动反馈

🗓 中期答辩活动

+ 申报基地

🏠 基地管理

+ 申报书屋

📖 书屋管理

+ 申报校外活动

📅 我的校外活动

+ 申报课题

📖 我的课题

已参加活动

-

搜索

活动名称	活动时间	认定时长	发布人	操作
传承家风，弘扬孝道——敬亲孝老实践活动	2026-01-27 08:00到 2026-02-25 08:00	8 小时	武宇航	查看 打印
追寻红色足迹，传承革命精神——铭记历史，凝心铸魂，强基固本	2026-01-26 08:00到 2026-02-19 08:00	20 小时	武宇航	查看 打印
2026年寒假母校回访专题实践	2026-01-01 08:00到 2026-03-31 08:00	19 小时	严景云	查看 打印
寻迹家乡：非遗传承与乡土文化焕新	2025-07-29 08:00到 2025-08-25 08:00	未认定时长	武宇航	查看
传承家风，弘扬孝道——孝老敬亲实践活动	2025-01-13 08:00到 2025-01-31 08:00	8 小时	武宇航	查看 打印
返乡乡 追梦启航——乡村振兴与家乡建设专项实践活动	2025-01-11 08:00到 2025-01-22 08:00	未认定时长	武宇航	查看
追寻红色足迹，传承革命精神——铭记历史，筑牢红色根基	2025-01-11 08:00到 2025-01-22 08:00	15 小时	武宇航	查看 打印
凝心铸魂，强基固本——五育融合助力成长	2025-01-11 08:00到 2025-01-21 08:00	未认定时长	武宇航	查看
2025年“梦圆南开 心系母校”的团体实践	2024-12-01 08:00到 2025-02-28 08:00	23 小时	张颖	查看 打印

🏠 基本资料

+ 申报活动

📅 我的活动

👤 已报名活动

🕒 已参加活动

📝 活动反馈

🗓 中期答辩活动

+ 申报基地

🏠 基地管理

+ 申报书屋

📖 书屋管理

+ 申报校外活动

📅 我的校外活动

我的校外活动

活动名称	活动时间	活动时长	活动状态	审核状态	操作
承袭古韵，赓续非遗——广府太极拳文化传承与推广	2026-08-01 08:00到 2026-08-09 12:00	22 小时	已结束	已通过	编辑
寻迹家乡：非遗传承与乡土文化焕新	2025-07-29 08:00到 2025-08-16 08:00	22 小时	已结束	已通过	编辑

大一寒假是否参加 否
社会实践：

大二寒假是否参加 是
社会实践：

大三寒假是否参加 是
社会实践：

大一暑假是否参加 否
社会实践：

大二暑假是否参加 是
社会实践：

大三暑假是否参加 是
社会实践：

(二) 志愿服务与社会实践（上限6分）

- 2、社会实践（上限为2分）
 - (1) 社会实践证明



-2025年“梦圆南开 心系母校”的团体实践-
社会实践证明

武宇航 同学：

于2024年12月01日08:00至2025年02月28日08:00参与了以2025年“梦圆南开 心系母校”的团体实践为主题的社会实践活动，社会实践时长为23小时，该社会实践活动已在共青团南开大学委员会社会实践和志愿服务督导中心处备案。

特此证明。



南开大学社会实践和志愿服务督导中心
2025年2月





追寻红色足迹，传承革命精神——铭记历史，筑牢红色根基-
社会实践证明

武宇航 同学：

于2025年01月11日08:00至2025年01月22日08:00参与了以追寻红色足迹，传承革命精神——铭记历史，筑牢红色根基为主题的社会实践活动，社会实践时长为15小时，该社会实践活动已在共青团南开大学委员会社会实践和志愿服务督导中心处备案。

特此证明。



南开大学社会实践和志愿服务督导中心
2025年1月





-传承家风，弘扬孝道——‘孝老敬亲’实践活动-
社会实践证明

武宇航 同学：

于2025年01月13日08:00至2025年01月31日08:00参与了以传承家风，弘扬孝道——‘孝老敬亲’实践活动为主题的社会实践活动，社会实践时长为8小时，该社会实践活动已在共青团南开大学委员会社会实践和志愿服务督导中心处备案。

特此证明。



南开大学社会实践和志愿服务督导中心
2025年1月



(二) 志愿服务与社会实践（上限6分）

- 2、社会实践（上限为2分）
 - (1) 社会实践证明



-2026年寒假母校回访专题实践-
社会实践证明

武宇航 同学:

于2026年01月01日08:00至2026年03月31日08:00参与了以2026年寒假母校回访专题实践为主题的社会实践活动，社会实践时长为19小时，该社会实践活动已在共青团南开大学委员会社会实践和志愿服务督导中心处备案。

特此证明。



南开大学社会实践和志愿服务督导中心
2026年3月



追寻红色足迹，传承革命精神——铭记历史，凝心铸魂，强基固本-
社会实践证明

武宇航 同学:

于2026年01月26日08:00至2026年02月19日08:00参与了以追寻红色足迹，传承革命精神——铭记历史，凝心铸魂，强基固本为主题的社会实践活动，社会实践时长为20小时，该社会实践活动已在共青团南开大学委员会社会实践和志愿服务督导中心处备案。

特此证明。



南开大学社会实践和志愿服务督导中心
2026年2月



-传承家风，弘扬孝道——‘敬亲孝老’实践活动-
社会实践证明

武宇航 同学:

于2026年01月27日08:00至2026年02月25日08:00参与了以传承家风，弘扬孝道——‘敬亲孝老’实践活动为主题的社会实践活动，社会实践时长为8小时，该社会实践活动已在共青团南开大学委员会社会实践和志愿服务督导中心处备案。

特此证明。



南开大学社会实践和志愿服务督导中心
2026年2月

(二) 志愿服务与社会实践（上限6分）

2、社会实践（上限为2分）

(2) 社会实践获荣誉称号

南开大学 2026 年“南风追梦”
优秀团队获奖证明

兹证明，石家庄精英中学团队于 2026 年寒假期间参加南开大学“南风追梦”母校回访寒假社会实践活动，实践过程中组织有序、表现优异。经学校综合评审，授予该团队南开大学 2026 年“南风追梦”母校回访寒假社会实践优秀团队三等奖荣誉称号。

团队成员名单如下：

职务	姓名	学号	职务	姓名	学号
队长	孙效涵	2412850	队员	赵浩博	2410161
队员	赵浩宇	2512640	队员	梁靖泽	2412852
队员	李艺瑶	2512647	队员	刘馨蔓	2412860
队员	王晴琳	2512639	队员	韩依诺	2412858
队员	魏天雨	2412859	队员	武宇航	2310137
队员	于子朔	2412854	队员	张珂豪	2510020
队员	尤浩轩	2512649	队员	米若天	2512636
队员	李锦皓	2512641	队员	邓烁	2512646
队员	何思瀚	2512635	队员	张森博	2412843
队员	田宇翔	2512645	队员	卢山	2412844
队员	侯雅馨	2512642	队员	谷保奇	2512648
队员	申一博	2412851	队员	张雅程	2412855

特此证明。

南开大学本科生招生办公室

2026 年 9 月 3 日

社会实践活动名称：2026年寒假母校回访专题实践

获奖时间：2026年6月1日

获奖名称：南开大学2026年“南风追梦”母校回访寒假社会实践先进队伍

是否为第一负责人：否

获奖等级：校级

(三) 艺体素质与技能特长（上限2分）

1、代表学校、学院参加艺术体育活动



参加活动名称： 南开大学2025年运动会

参加时间： 2025年10月12日

所属学院： 密码与网络安全学院

证明人： 梁昊

(三) 艺体素质与技能特长 (上限2分)

1、代表学校、学院参加艺术体育活动



参加活动名称： 南开大学2024年运动会

参加时间： 2024年9月21日

所属学院： 密码与网络安全学院

证明人： 方京伟

(三) 艺体素质与技能特长（上限2分）

2、参加艺术体育活动获奖情况



参加活动名称：	南开大学“五月的鲜花”合唱比赛	是否为第一负责人：	否
获奖名称：	南开大学“五月的鲜花”合唱比赛优秀奖	团队人数：	20人以上
获奖等级：	校级	所属学院：	计算机学院和密码与网络安全学院
获奖名次：	优胜奖	获奖时间：	2024年5月19日