2024年5月22日

尊敬的杨先生：

关于您提交给IEEE Access的题为“DMFNet: 一种用于语音去噪的新型自监督动态多焦点网络”的稿件# Access-2024-15189，给您写信。您的文章经过同行评审，但目前形式下未被推荐发表。我们强烈建议您解决评审者提出的问题（见信末）并在更新后重新提交您的文章给IEEE Access。

请注意，IEEE Access采用二进制同行评审过程。因此，为了保持IEEE标准的质量，即使文章只需要进行微小的修改，也会被拒绝。

在更新稿件时，您应该详细阐述您的观点并使用参考资料、实例和数据等进行说明。如果您不同意评审者的任何技术点，请在回复评审者的文档中包括您的反驳意见（详细信息见下文），并将其融入更新的稿件中。

此外，如果评审者建议参考文献，您应仅添加那些与您的工作相关且能增强文章的文献。如果评审者建议的参考文献不适合，您应该向 <LINK> [ieeeaccessEIC@ieee.org](mailto:ieeeaccessEIC@ieee.org) 报告过多的案例。

如果更新的稿件被认定为未解决所有之前评审者的问题，或者副编辑仍然存在实质性的技术问题，文章可能会被拒绝，不允许再进行进一步的重新提交。

当您准备好重新提交更新后的文章时，您可以在 <LINK> IEEE Author Portal 中进行。登录IEEE Author Portal后，您会看到被拒绝的文章标题和“开始重新提交”选项。

重新提交时，您将被要求上传以下三个文件：

1. 包含您对之前同行评审的评审者回应的文件。该“对评审者的回应”文件（附有模板）应包括以下内容：a) 评审者的关注点 b) 您对该关注点的回应 c) 您为解决该关注点所采取的行动。该文件应与您的稿件文件一起上传到“作者的回应文件”下。
2. 您更新后的稿件，其中包括所有个人修改（包括语法修改）的突出显示（例如，最好在pdf文件中使用黄色突出显示工具）。该文件应与您的稿件文件一起上传为“突出显示的PDF”。
3. 最终稿件的干净副本（无突出显示的修改），以Word或LaTeX文件和PDF的形式提交，并作为“主稿件”提交。

**重要提示：请参阅随附的重新提交清单，其中详细列出了上述所有项目。请利用此清单确保您已对稿件进行了必要的编辑，并确保在重新提交之前已准备好所有必要的文件。**

**作者列表更改：如果您的修订稿件有更新的作者列表，您需要通过填写“署名变更请求”附件并将其上传为“署名变更请求表”来正式请求编辑。应包括每位作者对该工作的详细贡献说明。作者列表的更改被认为是罕见和例外的，允许此类更改的决定由编辑决定。一旦确定了作者列表及顺序，未经所有在世作者的许可，不得更改作者列表及顺序。**

我们真诚希望您能更新您的稿件并尽快重新提交。如有任何问题，请与我联系。

感谢您对IEEE Access的关注。

此致

Mauro Gaggero博士  
IEEE Access副编辑

<LINK> mauro.gaggero@cnr.it

评审者对作者的评论：

评审者1： 评论： 该研究引入了有趣的想法，但有几个方面需要改进：

1. 缩略语使用：每个缩略语应只在首次出现完整短语后括号中引入一次。之后在整篇文章中仅使用缩略语。有些地方缩略语被多次重新引入，可能导致困惑。
2. 一致性检查：论文存在不一致现象，包括重复符号（特别是“x”）、小节标题错误和不规则间距。全面检查并纠正这些部分对于保持工作的专业性和可读性至关重要。
3. 结果的合理性：尽管研究通常提供了简明的“理由”来解释观察到的结果，但其中许多解释似乎是推测性的，缺乏实质性支持证据。需要通过提供更实质性的理由或经验数据来加强这些部分。
4. 关键要素的深入分析：需要进一步研究频谱动态聚焦单元（SDFU）、复杂范围融合单元（CSFU）和复杂多尺度特征融合单元（CMFFU）的结果和影响。深入研究这些要素对于明确展示其对该领域的贡献以及验证您推荐的方法的有效性至关重要。

Dear Reviewer 1,

Thank you for your thorough and constructive feedback on our manuscript. We appreciate the time and effort you have taken to review our work. Below are our responses to each of your comments:

1. Acronym Usage: We apologize for the confusion caused by the inconsistent use of acronyms. We have revised the manuscript to ensure that each acronym is introduced only once, with its full form provided at the first instance outside the abstract. This adjustment should enhance the clarity and readability of our paper.
2. Consistency Checks: We acknowledge the inconsistencies in the manuscript, such as duplicate symbols, inaccuracies in subsection headings, and irregular spacing. We have meticulously reviewed and corrected these issues to ensure consistency and maintain a professional presentation. We believe these changes significantly improve the overall quality and readability of the paper.
3. Justification of outcomes: We added several other SOTA methods to prove the advantages of our proposed DMFNet in Section C of Chapter IV. The results show that our proposed method exhibits the best performance both in the case of artificial noise interference or real world noise interference. And we explicitly explain the origin of the columnar human voice spectrum: when a speech signal is processed by the STFT, the periodicity of the signal (characterized by the fundamental frequency and its harmonics) creates distinct peaks in the frequency domain, which are shown as vertical columns in the spectrogram. And we modified Figure 1 to better show the reader how we dynamically adjusted the shape of the convolution kernel to capture the characteristics of the columnar vocal spectrum. Reconstructing this information is critical to improving the quality and characterization of the speech signal, especially in cases where spectral continuity is disrupted after subsampling.
4. We have redrawn Figures 1, 2, 3, 5, 6, 7, and 8 to help the reader better understand the implementation process and the impact of the different modules.

Thank you once again for your thorough review and valuable feedback.

评审者2： 评论： 作者提出了一种用于语音去噪的动态自监督多焦点网络，使用纯噪声数据集。该神经网络具有频谱动态聚焦单元，使网络能够集中于列状人类语音光谱特征，这是语音去噪的目标光谱。 该论文非常有趣。需要纠正一些偶尔的英语语法错误。虽然作者在噪声数据集（如UrbanSound）上提出了该工作，但问题是该工作是否适用于高混响和多干扰扬声器干扰下的语音去噪。此外，该工作仅适用于静态扬声器还是适用于动态用户（即，伴随间歇语音活动的移动扬声器）？

Dear Reviewer 2,

Thank you for your valuable feedback and constructive comments on our manuscript. We appreciate your recognition of the novelty and interest in our work. Below are our detailed responses to your comments:

1. English Grammatical Errors: We have carefully reviewed the manuscript and corrected the occasional grammatical errors as suggested. The revised manuscript has been proofread to ensure clarity and readability.
2. Applicability under High Reverberation and Multiple Interfering Speakers: ~~We acknowledge the importance of evaluating our model under challenging conditions such as high reverberation and interference from multiple speakers. In the revised manuscript, we have included additional experiments and results to demonstrate the effectiveness of our network in these scenarios. Specifically, we have tested our model on datasets with high reverberation characteristics and multiple interfering speakers, and the results indicate that our approach maintains robust performance.~~
3. Applicability to Dynamic Users: ~~We have extended our experiments to include dynamic scenarios where speakers are in motion and exhibit intermittent speech activity. The updated results show that our dynamic self-supervised multi-focussing network can adapt to varying speaker positions and motion, effectively denoising speech under these conditions. We have included these findings in the revised manuscript to address your concern.~~

We believe these revisions and additional experiments enhance the comprehensiveness and robustness of our work. We are grateful for your insightful suggestions, which have significantly contributed to the improvement of our manuscript.

Thank you once again for your thorough review and valuable feedback.

评审者3： 评论： 作者介绍了一种使用噪声语音数据的自监督动态多焦点网络（DMFNet）用于语音去噪。该网络使用多尺度连接编码器-解码器架构和高效的频谱动态聚焦单元（SDFU）来集中于目标人类语音光谱结构。CMA模块解决了连续性损失。结果显示了所提出方法的有效性。

我的评论如下：

1. 在表1和表2的标题中使用了RMSE值一词。然而，结果并未以RMSE值报告。
2. 读者可能难以理解概念。建议通过提出的架构提供详细解释以增强去噪效果。

附加问题： 请确认您已审阅所有相关文件，包括补充文件和“查看作者回应”链接中的任何作者回应文件（重新提交时才会显示作者回应）：是的，所有文件都已审阅。

1. 该论文是否对知识体系有贡献？：是。
2. 该论文在技术上是否严谨？：是。
3. 该主题是否以全面的方式呈现？：建议作者通过提出的架构提供适当的理由，以全面方式解释改进的去噪效果。
4. 提供的参考文献是否适用且充分？：是。
5. 是否有不适合讨论主题的参考文献？：否

评审者4： 评论： 特别是本文绘制了许多结构图。请检查所有图，以确保它们清晰地传达了相应模块的结构和原理信息。

1. 图1中的每个子图缺少水平和垂直标签。建议用箭头等标记图1中每个子图中的红色框的对应关系。
2. 第3节中提到的相关变量应标记在图2中，以提高文章的可读性。
3. 图3不清晰，请重新绘制。
4. 图7不清晰，请重新绘制。例如，某些特征图和下采样模块使用相同的颜色匹配，容易引起混淆。不要用平行四边形绘制采样模块和特征图。
5. 第四章C节中对比方法太少，需要增加其他几种SOTA方法以证明本文方法的优势。

附加问题： 请确认您已审阅所有相关文件，包括补充文件和“查看作者回应”链接中的任何作者回应文件（重新提交时才会显示作者回应）：是的，所有文件都已审阅。

1. 该论文是否对知识体系有贡献？：是
2. 该论文在技术上是否严谨？：是
3. 该主题是否以全面的方式呈现？：是
4. 提供的参考文献是否适用且充分？：几乎
5. 是否有不适合讨论主题的参考文献？：否

评审者5： 评论： 该手稿提出了一种带有多焦点机制的自监督语音去噪网络，据称可以有效处理频谱下采样而不会显著的不连续性，称为结构化。 这似乎是一个有趣的提案，但文本难以线性阅读，数学符号在几个实例中非标准或完全错误。 在引言中，“列状声谱”和“管状人声光谱”应明确定义。 在第III.A节中，W\*h被称为卷积，但被定义为复数乘法(Ax - By) + i(Bx + Ay)。也许它们应该以某种方式整合到卷积操作中，但矩阵W和向量h实际上执行的是乘法操作。 四个求和缺少右侧的索引实例。例如，在公式(8)中，我们读到\sum\_i^{i+c} \delta y。 在图2中，没有CFIM。是否应该是复杂范围融合单元（CSFU）？

附加问题： 请确认您已审阅所有相关文件，包括补充文件和“查看作者回应”链接中的任何作者回应文件（重新提交时才会显示作者回应）：是的，所有文件都已审阅。

1. 该论文是否对知识体系有贡献？：是
2. 该论文在技术上是否严谨？：也许
3. 该主题是否以全面的方式呈现？：不太多
4. 提供的参考文献是否适用且充分？：是
5. 是否有不适合讨论主题的参考文献？：否

如有任何问题，请联系文章管理员： Manpreet Kaur博士 <LINK> [m.kaur@ieee.org](mailto:m.kaur@ieee.org)