arXiv:2301.04320v1 [cs.SD] 11 Jan 2023

重新审视用于单声道语音增强的复值深度神经网络

Haibin Wu、*Ke Tan*、*Buye Xu*、*Anurag Kumar*、*Daniel Wong*

Meta Reality Labs Research，美国

摘要

尽管人们已多次尝试采用复值深度神经网络（DNNs），但对于单声道语音增强而言，复值深度神经网络是否普遍比实值深度神经网络更有效，仍是一个尚无定论的问题。本研究旨在通过系统地将复值深度神经网络与其实值对应模型进行比较，给出批判性评估。具体来说，我们考察复值深度神经网络的基本单元，包括线性层、卷积层、长短期记忆网络（LSTM）和门控线性单元。通过比较近期提出的门控卷积循环网络（GCRN）中基本构建模块的复值与实值版本，我们展示了基本模块的不同机制如何影响性能。我们还发现，当模型较小时，采用复值运算会限制模型容量。此外，我们还考察了近期提出的两种复值深度神经网络，即深度复值卷积循环网络（DCCRN）和深度复值 U-Net（DCUNET）。评估结果表明，这两种深度神经网络的性能均与其实值对应模型相当，但需要更多计算。基于这些全面的比较，我们得出结论：对于单声道语音增强，复值深度神经网络并不能比其实值对应模型带来性能提升，因此由于其计算成本更高，并不那么理想。

索引词— 单声道语音增强、复值神经网络、计算成本、深度学习

1. 引言

近年来，鉴于相位对语音质量的重要性 [1–9]，单声道语音增强模型在复数域取得了令人瞩目的性能提升。近期一项研究 [10] 提出了复值深度神经网络的关键基本组件，并指出复值参数从计算、生物学和信号处理的角度来看具有多种优势。复值深度神经网络采用复值算术运算，似乎适用于复数域语音增强，因为在这种任务中，深度神经网络需要学习复数频谱图。受此直觉启发，已有多项研究 [3, 7, 11–16]将复值深度神经网络用于单声道语音增强。然而，据我们所知，尚无研究证明，在网络结构和模型大小相同的情况下，复值深度神经网络的性能优于其实值对应模型。Drude 等人 [17]比较了用于波束成形的全连接层实值和复值深度神经网络，发现复值深度神经网络的计算开销更大，性能却并不优于实值深度神经网络。对于单声道语音增强，尽管近期复值深度神经网络带来了令人瞩目的性能提升，但其优势是否从根本上源于复值特性，仍不清楚。

一项近期颇受关注的模型 DCCRN [7]通过将卷积层和 LSTM 层替换为复值对应层，扩展了 [18] 中的卷积循环网络，以估计理想复数比率掩码。DCCRN 的性能优于早期研究中的模型，这使学界开始关注复值深度神经网络在语音增强中的有效性。然而，我们认为，由于缺乏 DCCRN 与其实值对应模型之间的系统比较，现在就将性能提升归因于复值运算还为时过早。在其实值对应模型中，仅将复值层替换为相应的实值层，而其他所有配置均保持不变，包括输入特征、训练目标、训练目标函数、网络结构和模型大小。缺少这种同等条件下的比较，就难以证实复值深度神经网络确实带来了性能提升。

本研究通过全面比较，系统考察复值深度神经网络与其实值对应模型，从而进行批判性评估：

1). 基于复值计算原理 [10], ，我们构建并研究了复值深度神经网络的基本单元，包括线性层、卷积层/反卷积层、长短期记忆网络和门控线性单元，并比较了它们在单声道语音增强任务中的表现与其实值对应模型的表现。
2). 为全面研究不同层拓扑中的复值运算，我们采用 GCRN——一种最初为复数域语音增强开发的实值深度神经网络，它整合了多种层类型。我们列举了 GCRN 基本构建模块的所有不同版本，并展示了基本模块中不同计算机制对性能的影响。我们发现，包含复值组件的模型并未优于实值对应模型。此外，鉴于许多现实应用需要计算高效模型，我们在模型大小极小的设置下进行了同样的比较。我们发现，在这种设置下，与实值运算相比，复值运算甚至会降低语音增强性能。

3）。近期有两种基于复数运算的出色模型 DCCRN [7] 和 DCUNET [3], ，在单声道语音增强方面展现出良好的性能。本文评估了它们参数数量相同的实值版本，并从增强性能和训练稳定性两个方面，考察了不同损失函数、学习率和小批量大小的影响。实验结果表明，复值版本的性能并不优于其实值对应模型，同时计算成本更高。这与 [19] 中的观察结果一致。

这项工 作是在 H. Wu 作为研究科学家实习生任职于 M eta.

2. 方法论

本节介绍复值深度神经网络的基本构建模块，随后介绍案例研究设计。

2.1. 构建模块

2.1.1. 线性

全连接层、卷积层和反卷积层均由矩阵乘法构成。为简化描述，我们省略偏置。将输入复值特征矩阵记为$X = X_r + jX_i$，复值参数矩阵记为 $W = W_r + jW_i$，则复值输出可表示为:

$$Y = (X_r W_r - X_i W_i) + j(X_r W_i + X_i W_r), \tag{1}$$

其中，Y表示复值层的输出特征，下标 r 和 i 分别表示实部和虚部。

2.1.2. 激活函数

给定一个复值表示 z，激活函数分别作用于实部和虚部:

$$a = f(Re\ z) + jf(Im\ z), \tag{2}$$

其中，a 是输出表示，Re 和 Im 分别提取实部和虚部，f 表示激活函数。

2.1.3. 长短期记忆网络

对于 LSTM 层，我们有两个版本:

准复值长短期记忆网络 在 [7], 中，复值 LSTM 运算被视为分别对实部和虚部进行的两个独立运算。具体而言，它们初始化两个实值子 LSTM 层，即 LSTM_r 和 LSTM_i，分别对应实部和虚部 LSTM。给定输入特征 $X = X_r + jX_i$，输出特征可推导如下:

$$\begin{aligned} &F_{rr} = \text{LSTM}_r(X_r), F_{ir} = \text{LSTM}_r(X_i), \\ &F_{ri} = \text{LSTM}_i(X_r), F_{ii} = \text{LSTM}_i(X_i), \\ &F_{out} = (F_{rr} - F_{ii}) + j(F_{ri} + F_{ir}), \end{aligned} \tag{3}$$

其中，F_{out} 是输出特征。

全复数值LSTM除了不会在子长短期记忆网络层内执行复值运算的准复数值LSTM之外，我们还研究了全复数值LSTM，它完全遵循复数的运算规则。该LSTM中的每次矩阵乘法和激活函数都严格遵循第 2.1.1 节和第 2.1.2 节中的运算规则。

2.1.4. 门控线性单元

门控线性单元 [20] 是一种广泛使用的层拓扑结构，由两个分离的卷积层和一个门控操作组成。两个分离的卷积层处理相同的输入，并分别产生输出 $F^{(1)}$ 和 $F^{(2)}$。对 $F^{(2)}$ 应用 Sigmoid 函数以导出一个门控值，然后将其与 $F^{(1)}$ 逐元素相乘，得到门控线性单元的输出。在复值门控线性单元中，设 $F^{(1)} = F_r^{(1)} + jF_{(1)\,i}$和$F_{(2)} = F_{(2)r} + jF_{(2)\,i}$为两个卷积层的输出。我们推导出两种门控机制，即分离门控和幅度门控。

分离门控 对于分离门控，我们分别对 $F^{(2)}$ 的实部和虚部应用 Sigmoid 函数，这相当于一个复值门。该门的实部和虚部分别与 $F_r^{(1)}$和 $F_i^{(1)}$ 逐元素相乘。

幅度门控与分离门控不同，幅度门控根据复值特征图 $F^{(2)}$的幅值计算实值门 $F^{(g)}$:

$$F^{(g)} = (\sigma(|F^{(2)}|) - 0.5) \times 2, \tag{4}$$

其中，σ 表示 Sigmoid 函数，$|\cdot|$ 用于提取复值特征图的幅值。由于幅值非负，对其应用 Sigmoid 函数后，所得数值始终介于 0.5 和 1 之间。因此，我们使用仿射变换将门控值归一化到 0 至 1 的范围。所得门控同时应用于 $F^{(1)}$的实部和虚部。这种幅度门控能够保留 $F^{(1)}$ [21]的相位。

2.2. 案例研究设计

本节将仔细设计案例研究，并详细阐述每项案例研究的依据和目标。在这些案例研究中，所有用于比较的实值模型和复值模型在配置上均保持一致，包括输入特征、训练目标、训练目标函数、网络结构和模型大小。

基本单元 本案例研究比较第 2.1 节定义的不同复值层及其对应的实值层，考察它们的增强性能和计算成本。具体而言，我们比较：1）由三个复值线性层堆叠而成的模型及其对应的实值模型；复值模型的两个隐藏层各有 406 个单元，实值模型的两个隐藏层各有 512 个单元。这样的配置可确保两个模型的参数数量几乎相同。请注意，每个隐藏层后均接一个修正线性单元函数；2）准复值长短期记忆网络、全复数值LSTM和实值长短期记忆网络，这三种模型均包含三个长短期记忆网络层，后接一个线性输出层。这三种模型的每个长短期记忆网络层分别包含 732、732 和 1024 个单元。准复值长短期记忆网络和全复数值LSTM采用第 2.1.3 节所述的实现方式；3）DCUNET（一种卷积编码器-解码器模型）及其在 [3], 中提出的实值对应模型（RUNET）。在 RUNET 中，所有复值卷积层、反卷积层和线性层均替换为对应的实值层。与 1）和 2）类似，我们对 RUNET 的超参数（例如卷积层的输出通道数）稍作调整，使其模型大小与 DCUNET 几乎相同。请注意，所有这些模型的训练目标均为学习复数频谱映射。

GCRN GCRN [5]是我们研究的代表性模型，因为它由多种类型的层组成，包括卷积层/反卷积层、门控线性单元、LSTM 层和线性层。原始 GCRN 有两个解码器，一个用于实部估计，另一个用于虚部估计。我们改为使用一个共享解码器来处理实部和虚部，这对应于解码器最后一个反卷积层中的两个输出通道。通过将每一层替换为其复值对应层，可以自然地将这种架构转换为复值版本进行比较。在本案例研究中，我们旨在探究：1）将 GCRN 的特定层替换为其复值对应层是否能够带来更好的性能；2）当模型的参数量受到相对严格的限制时，使用复值运算会如何影响语音增强性能；3）哪种门控机制在

表1：不同基本单元的研究，其中乘加（MAC）运算次数以1秒的信号为基准进行测量 al.

	SNR	含噪	(1a).C-LSTM	(1b).准 C-LSTM	(1c).长短期记忆网络	(1d).C-Linear	(1e).R-Linear	(1f).DCUNET	(1g).RUNET
STOI	-5 分贝	0.69	0.85	0.86	0.86	0.61	0.61	0.85	0.85
	0 dB	0.78	0.90	0.91	0.91	0.70	0.70	0.90	0.90
	5 dB	0.85	0.94	0.94	0.94	0.76	0.76	0.94	0.94
宽带感知语音质量评估	-5 分贝	1.11	1.65	1.71	1.69	1.12	1.12	1.64	1.70
	0 dB	1.15	1.95	2.02	2.00	1.17	1.28	1.92	2.00
	5 dB	1.24	2.29	2.35	2.34	1.24	1.25	2.27	2.36
尺度不变信号失真比（分贝）	-5 分贝	-5.00	10.80	11.10	10.87	0.92	1.23	10.80	10.87
	0 dB	0.05	13.62	13.92	13.78	4.69	4.94	13.79	13.86
	5 dB	5.01	16.36	16.64	16.55	7.19	7.57	16.74	16.84
# 参数	-	-	23.35 M	23.35 M	23.62 M	0.59 M	0.59 M	3.10 M	3.12 M
# 乘加运算次数	-	-	5.90 G	5.90 G	2.98 G	119.59 M	59.88 M	56.69 G	19.87 G

表 2：对 GCRN 中不同复值组件的研究。♣、 ◇、 ♡ 分别表示在瓶颈层中使用准复值长短期记忆网络、复值卷积层和复值反卷积层。“- Separate”和“-Magnitude”分别表示在门控线性单元中使用分离门控机制和幅值门控机制； ⊙ 表示模型采用复数比率掩蔽，而非 [5] 中原先使用的复数频谱映射。

		STOI			宽带感知语音质量评估			尺度不变信号失真比（分贝）			# 参数	# 乘加运算次数
		-5 分贝	0 dB	5 dB	-5 分贝	0 dB	5 dB	-5 分贝	0 dB	5 dB		
	含噪	0.69	0.78	0.85	1.11	1.15	1.24	-5	0.01	5.01	-	-
(2a)	GCRN（实值模型）	0.84	0.90	0.94	1.57	1.87	2.24	8.30	11.29	14.13	9.25 M	1.72 G
(2b)	GCRN + ♣	0.83	0.90	0.94	1.55	1.85	2.22	8.22	11.17	13.98	9.25 M	2.57 G
(2c)	GCRN + ◇- 分离	0.83	0.90	0.93	1.53	1.80	2.15	7.64	10.51	13.22	9.12 M	1.72 G
(2d)	GCRN + ◇- 幅值	0.83	0.90	0.94	1.56	1.85	2.23	7.66	10.63	13.43	9.12 M	1.72 G
(2e)	GCRN + ♣ + ◇ - 分离	0.83	0.90	0.93	1.52	1.81	2.16	8.14	11.16	14.02	9.12 M	2.57 G
(2f)	GCRN + ♣ + ◇ - 幅值	0.84	0.90	0.94	1.56	1.87	2.24	7.89	10.89	13.76	9.12 M	2.57 G
(2g)	GCRN + ◇ + ♡ - 分离	0.83	0.89	0.93	1.53	1.83	2.20	7.95	10.96	13.88	8.83 M	1.72 G
(2h)	GCRN + ◇ + ♡ - 幅值	0.83	0.90	0.94	1.54	1.85	2.23	7.67	10.75	13.79	8.83 M	1.72 G
(2i)	GCRN + ♣ + ◇ + ♡ - 分离	0.82	0.89	0.93	1.52	1.80	2.15	7.62	10.70	13.52	8.83 M	2.57 G
(2j)	GCRN + ♣ + ◇ + ♡ - 幅值	0.83	0.90	0.94	1.57	1.88	2.27	7.65	10.87	13.82	8.83 M	2.57 G
(2A)	GCRN ⊙（实值模型）	0.83	0.89	0.93	1.50	1.79	2.16	7.28	10.33	13.40	9.25 M	1.72 G
(2J)	GCRN + ♣ + ◇ + ♡ - 幅值 ⊙	0.82	0.89	0.93	1.47	1.74	2.10	7.25	10.33	13.46	8.83 M	2.57 G

表 3：GCRN 与 CGCRN 在模型大小相对较小时的比较。下标“M”和“S”分别表示中等大小和小型。

	SNR	含噪	CGCRNM	GCRNM	CGCRNS	GCRNS
STOI	-5 分贝	0.69	0.81	0.81	0.75	0.79
	0 dB	0.78	0.88	0.88	0.83	0.86
	5 dB	0.85	0.92	0.92	0.88	0.91
宽带感知语音质量评估	-5 分贝	1.11	1.39	1.42	1.29	1.35
	0 dB	1.15	1.62	1.64	1.47	1.54
	5 dB	1.24	1.93	1.95	1.71	1.82
尺度不变信号失真比（分贝）	-5 分贝	-5.00	6.60	7.25	4.14	5.82
	0 dB	0.05	9.75	10.26	6.80	8.94
	5 dB	5.01	12.58	12.94	8.73	11.85
# 参数（M）	-	-	2.26	2.36	0.61	0.63
# 乘加运算次数（M）	-	-	657.90	439.03	172.12	115.99

第 2.1.4 节中介绍的门控机制是更好的选择，这需要从训练稳定性和增强性能两方面进行考量。请注意，针对 GCRN 中的瓶颈 LSTM，我们采用准复值长短期记忆网络进行研究。

DCCRN 在 [7], 中，作者将 DCCRN 所取得的性能提升归因于复数乘法约束，他们认为该约束有助于深度神经网络更有效地学习复数表示。然而，他们并未在相同配置下将 DCCRN 与其实值对应模型进行比较。因此，很难证明性能提升确实源于这一因素；提升可能是由复值运算或模型设计中的其他组件带来的。本案例研究旨在探究 DCCRN 是否能在参数量相同的情况下优于其实值对应模型。具体而言，我们采用“DCCRN-E”配置，该配置在 [7]中取得了最佳性能。为得到相应的实值版本，我们只需将复值层替换为其实值对应层，并减少编码器中的通道数至 [32, 64, 64, 64, 128, 256] ，

以保持参数量不变。

3. 实验

3.1. 实验设置

在我们的实验中，我们使用 Interspeech2020 DNS 挑战赛训练语音数据集 [22] 来构建训练集、验证集和测试集，其中包含约 65000 条语音信号，共由 1948 位说话人发出。我们将这些说话人随机划分为三个互不重叠的集合，分别用于训练、验证和测试，包含 1753 位（∼90%）、97 位（∼5%）和 98 位（∼5%）说话人。同样，我们将包含约 65000 条信号的 DNS 挑战赛噪声数据集按 90%、5% 和 5% 的比例划分为训练集、验证集和测试集。通过随机配对语音和噪声信号，我们构建了一个包含 500000 条含噪混合语音的训练集，以及一个包含 1000 条含噪混合语音的验证集；这两个集合中的信噪比（SNR）均在 -5 至 5 分贝之间随机采样。按照相同的流程，我们在不同信噪比下构建了三个测试集，即 -5、0 和 5 分贝。请注意，所有语音和噪声信号在混合前都会随机截取为 10 秒。我们还使用 DNS 挑战赛发布的合成测试集进行评估。

所有信号均以 16 千赫采样。通过短时傅里叶变换获取频谱图。我们采用 Adam 优化器训练所有模型。采用多种指标衡量语音增强性能，包括宽带感知语音质量评估 (WB-PESQ) [23], 短时客观可懂度 (STOI) [24], 尺度不变信号失真比 (SI-SDR) [25], DNSMOS P. 835 [26] 和 NORESQA-MOS [27]。

3.2. 实验结果

基本单元在表1中，1). 第(1a)、(1b)和(1c)列分别表示全复数值 LSTM、准复值长短期记忆网络和实值长短期记忆网络。2). 实值长短期记忆网络的乘加运算次数仅为两种复数值LSTM的一半。在这三个模型中，准复值长短期记忆网络的性能最佳，但相较于实值长短期记忆网络，其提升幅度很小。第(1d)列

表 4：采用不同训练目标函数的 DCCRN 实值版本与复值版本对比。“-Real”表示 DCCRN 的实值版本。“-SISDR”、“-L_1”和“-MSE”分别表示训练时使用尺度不变信号失真比、L_1 和均方误差（MSE）损失，其中 L_1 和 MSE 损失均在干净和估计的实部、虚部及幅值频谱图上计算。

	SNR	含噪	DCCRN-SISDR	DCCRN-Real-SISDR	DCCRN-L_1	DCCRN-Real-L_1	DCCRN-MSE	DCCRN-Real-MSE
STOI	-5 分贝	0.69	0.87	0.87	0.86	0.86	0.85	0.85
	0 dB	0.78	0.92	0.92	0.91	0.91	0.90	0.90
	5 dB	0.85	0.95	0.95	0.95	0.95	0.94	0.94
宽带感知语音质量评估	-5 分贝	1.11	1.78	1.80	1.73	1.69	1.55	1.56
	0 dB	1.15	2.13	2.14	2.05	2.00	1.83	1.86
	5 dB	1.24	2.51	2.54	2.43	2.38	2.16	2.19
尺度不变信号失真比（分贝）	-5 分贝	-5.00	11.01	11.06	8.36	8.28	8.09	8.18
	0 dB	0.05	14.00	14.06	11.25	11.19	11.20	11.27
	5 dB	5.01	16.99	17.05	14.36	14.27	14.41	14.54

表 5：在无混响的 DNS 挑战赛合成测试集上，采用不同训练目标函数的 DCCRN 实数值版本与复值版本的比较。

	含噪	DCCRN-SISDR	DCCRN-Real-SISDR	DCCRN-L_1	DCCRN-Real-L_1	DCCRN-MSE	DCCRN-Real-MSE
STOI	0.92	0.97	0.97	0.97	0.97	0.97	0.97
宽带感知语音质量评估	1.58	2.92	2.89	2.92	2.86	2.61	2.64
尺度不变信号失真比（分贝）	9.23	19.60	19.54	17.11	17.13	17.33	17.55
DNSMOS（OVRL）	2.48	3.30	3.33	3.28	3.30	3.19	3.20
NORESQA-MOS	1.90	4.31	4.34	4.27	4.31	3.80	3.96
# 参数	-	3.67 M	3.64 M	3.67 M	3.64 M	3.67 M	3.64 M
# 乘加运算次数	-	14.38 G	4.84 G	14.38 G	4.84 G	14.38 G	4.84 G

和第(1e)列分别表示由线性层构成的复数值和实值深度神经网络。尽管实值深度神经网络的乘加运算次数仅为复数值深度神经网络的一半，但其性能仍略胜后者。3). 第(1f)和第(1g)分别表示DCUNET及其对应的实值版本。我们发现，在增强性能和计算效率方面，实值UNET均优于DCUNET。

GCRN表 2 中，(2a) 是原始的实值 GCRN。(2b)-(2j) 是将部分组件替换为相应复值版本的模型。此外，(2A) 和 (2J) 的模型结构分别与 (2a) 和 (2j) 相同，但训练目标是复数比率掩蔽，而非复数频谱映射。表 3 中，我们将模型大小缩减至约 2 M 和 0.6 M，其中“CGCRN”表示与 (2j) 相同的配置。我们可以观察到：1). 将 GCRN 的组件替换为复值版本并未带来性能提升，如 (2a)-(2j) 所示。2). 在针对复数比率掩蔽进行训练的模型 (2A) 和 (2J) 的比较中，实值模型的表现略优于复值模型。3). 尽管幅度门控和分离门控的性能相近，但前者的训练损失曲线比后者稳定得多。这可能是因为幅度门控保留了相位信息，有助于稳定训练。4). 在小型模型设置中，实值模型的表现始终优于相应的复值模型。此外，随着模型大小减小，两者的性能差距会扩大。

DCCRN 表 4 和表 5 分别比较了 DCCRN 与其实值对应模型在我们的模拟测试集和 DNS 挑战赛合成测试集上的表现。得出以下观察结果：1). 对于三种不同的训练目标函数，即 SI-SDR、L_1 和均方误差，两种实值模型和复值模型在两个数据集的所有指标上表现几乎相同。以表 4 中采用 SI-SDR 训练损失的 -5 分贝情形为例，复值模型在短时客观可懂度、宽带感知语音质量评估和尺度不变信号失真比方面，相较于含噪混合语音的提升分别为 0.18、0.67 和 16.01 分贝；实值模型的相应提升分别为 0.18、0.69 和 16.06 分贝。2) 如表 5 所示，实值模型在 DNSMOS 和 NORESQA-MOS 两项指标上的得分略高于复值模型，即这两项指标与主观质量评分高度相关。3). 我们还比较了采用不同学习率和小批量大小的设置。我们发现，DCCRN 对不同学习率的适应性不如其实值对应模型。此外，两种模型在不同小批量大小下的表现非常相似。不过，由于篇幅限制，我们未展示这些比较结果。4). 实值模型的乘加运算次数仅为复值模型的三分之一。具体而言，复值模型的乘加运算次数为 14.38 G，而实值模型仅为 4.84 G。鉴于两种模型的表现几乎相同，复值模型在实际应用中的效率较低。

4. 总结

通过大量实验，我们针对单声道语音增强得出以下结论：1). 在参数数量相同的情况下，复值深度神经网络的性能与其实值对应模型相近。2). 当模型大小相对较小时，使用复值运算会损害增强性能。3). DCCRN 和 DCUNET 所取得的性能提升并非归因于复值运算的使用。此外，复值深度神经网络所需的乘加运算次数多于其实值对应模型，却没有带来任何性能提升。

一次复数乘法可以拆分为四次实数乘法。基于系统性的比较，我们认为，在模型大小和网络结构相同的情况下，实值深度神经网络有能力达到与其复值对应层相当的性能。尽管直觉上，相较于实值深度神经网络，复值深度神经网络似乎更适合处理复数频谱，但其计算成本更高，因此对于对效率敏感的实际应用而言并非更优选择。我们认为，目前没有充分证据证明复值深度神经网络在单声道语音增强方面优于实值深度神经网络。本研究表明，重新审视语音增强系统中复值运算的有效性绝非易事。

5. 参考文献

[1] D. S. Williamson、Y. Wang 和 D. L. Wang，“用于单通道语音分离的复数比率掩蔽”，*IEEE/ACM* 音频、语音与语言处理汇刊，第 24 卷，第 3 期，第 483–492 页，2015年。[2] S.-W. Fu、T.-y. Hu、Y. Tsao 和 X. Lu，“通过具有多指标学习的卷积神经网络进行复数谱图增强”，载于*IEEE*第*27*届信号处理机器学习国际研讨会。IEEE，2017年，第 1–6 页。[3] H.-S. Choi、J.-H. Kim、J. Huh、A. Kim、J.-W. Ha 和 K. Lee，“采用深度复值 U-Net 的相位感知语音增强”，*arXiv* 预印本 *arXiv:1903.03107*，2019年。[4] K. Tan 和 D. L. Wang，“采用卷积循环网络进行复数频谱映射以实现单声道语音增强”，载于*IEEE*声学、语音与信号处理国际会议。IEEE，2019年，第 6865–6869 页。[5] K. Tan 和 D. L. Wang，“利用门控卷积循环网络学习复数频谱映射以实现单声道语音增强”，*IEEE/ACM* 音频、语音与语言处理汇刊，第 28 卷，第 380–390 页，2019年。[6] Z. Ouyang、H. Yu、W.-P. Zhu 和 B. Champagne，“用于语音增强中复数谱图处理的全卷积神经网络”，载于*IEEE*声学、语音与信号处理国际会议。IEEE，2019年，第 5756–5760 页。

[7] Y. Hu, Y. Liu, S. Lv, M. Xing, S. Zhang, Y. Fu, J. Wu, B. Zhang 和 L. Xie，“DCCRN：用于相位感知语音增强的深度复数卷积循环网络”，载于*Interspeech*，2020年，第2472–2476页。

[8] R. Xu、R. Wu、Y. Ishiwaka、C. Vondrick 和 C. Zheng，“聆听寂静之声以实现语音降噪”，神经信息处理系统进展，第33卷，第9633–9648页，2020年。

[9] A. Li, W. Liu, X. Luo, G. Yu, C. Zheng, 和 X. Li，“两个人脑智慧胜过一个：一种用于单声道语音增强的两阶段复数频谱映射方法”，*IEEE/ACM* 音频、语音与语言处理汇刊，第 29 卷，第 1829–1843 页，2021。

[10] C. Trabelsi, O. Bilaniuk, Y. Zhang, S. Mehri, N. Rostamzadeh, S. Dmitriy Serdyuk, J. F. Santos, Y. Bengio, 和 C. J. Pal，《深度复数网络》，载于第*6*届国际学习表征会议，2018年。

[11] S. Lv、Y. Hu、S. Zhang 和 L. Xie，“用于语音增强的带信噪比估计的逐通道子带 DCCRN+”，发表于*Interspeech*，2021，页码 2816–2820。[12] S. Lv、Y. Fu、M. Xing、J. Sun、L. Xie、J. Huang、Y. Wang 和 T. Yu，“用于语音增强的具有可学习复数特征的超宽带 DCCRN（S-DCCRN）”，发表于*IEEE* 声学、语音与信号处理国际会议。IEEE，2022年，页码 7767–7771。[13] S. Zhao、T. H. Nguyen 和 B. Ma，“采用复数卷积块注意力模块和联合时频损失的单通道语音增强”，发表于*IEEE*声学、语音与信号处理国际会议。IEEE，2021，页码 6648–6652。

[14] Y. Sun、L. Yang、H. Zhu 和 J. Hao，“用于相位感知语音增强的漏斗式深度复数 U-Net”，发表于*Interspeech*，2021，页码 161–165。[15] K. N. Watcharasupat、T. N. T. Nguyen、W.-S. Gan、S. Zhao 和 B. Ma，“用于声学回声消除与噪声抑制的端到端复值多膨胀卷积神经网络”，发表于*IEEE*声学、语音与信号处理国际会议。IEEE，2022年，页码 656–660。

[16] H. Zhou, J. Zhu, “DCTCN：用于实时语音增强的深度复值时序卷积网络”，载于第*11*届智能控制与信息处理国际会议。IEEE，2021，第112–118页。

[17] L. Drude, J. Heymann, R. Haeb-Umbach, “复值神经网络用于语音增强是否恰当。”，载于*Interspeech*，2016，第1745–1749页。

[18] K. Tan, J. Chen, D. Wang, “用于实时语音增强的卷积循环神经网络。”，载于*Interspeech*，2018年，第2018卷，第3229–3233页。

[19] A. Pandey和D. L. Wang，“探索用于复数谱图增强的深度复数网络”，载于*IEEE*声学、语音与信号处理国际会议。IEEE，2019年，第6885–6889页。[20] Y. N. Dauphin、A. Fan、M. Auli和D. Grangier，“使用门控卷积网络进行语言建模”，载于国际机器学习会议。PMLR，2017年，第933–941页。[21] D. Hayakawa、T. Masuko和H. Fujimura，“将复值神经网络应用于语音识别的声学建模”，载于亚太信号与信息处理协会年度峰会暨会议。IEEE，2018年，第1725–1731页。[22] C. K. Reddy、V. Gopal、R. Cutler、E. Beyrami、R. Cheng、H. Dubey、S. Matusevych、R. Aichner、A. Aazami、S. Braun等，“Interspeech 2020深度噪声抑制挑战赛：数据集、主观测试框架和挑战赛结果”，载于*Interspeech*，2020年，第2492–2496页。[23] I. Rec，“P. 862.2：用于评估宽带电话网络和语音编解码器的建议书 P. 862 宽带扩展”，国际电信联盟，瑞士日内瓦，2005年。[24] C. H. Taal、R. C. Hendriks、R. Heusdens和J. Jensen，“一种用于预测时频加权噪声语音可懂度的算法”，*IEEE*《音频、语音与语言处理汇刊》，第19卷，第7期，第2125–2136页，2011年。[25] J. Le Roux、S. Wisdom、H. Erdogan和J. R. Hershey，“信号失真比——半生不熟还是恰到好处？”，载于*IEEE*声学、语音与信号处理国际会议。IEEE，2019年，第626–630页。[26] C. K. Reddy、V. Gopal和R. Cutler，“DNSMOS P. 835：一种用于评估噪声抑制器的非侵入式感知客观语音质量指标”，载于*IEEE*声学、语音与信号处理国际会议。IEEE，2022年，第886–890页。[27] P. Manocha和A. Kumar，“通过平均意见分并使用非匹配参考进行语音质量评估”，载于*Interspeech*，2022年，第654–658页。