arXiv:2303.11860v1 [cs.NE] 21 Mar 2023

采用脉冲神经元的在线 Transformer，实现快速假肢手控制

N. Leroux
彼得·格林贝格研究所、于利希研究中心，德国亚琛
n.leroux@fz-juelich.de

J. Finkbeiner
彼得·格林贝格研究所、于利希研究中心、亚琛工业大学，德国
j.finkbeiner@fz-juelich.de

E. Neftci彼得·格林贝格研究所、于利希研究中心、亚琛工业大学，德国
e.neftci@fz-juelich.de

摘要

Transformer 是大多数序列处理任务中的最先进网络。然而，Transformer 常用的自注意力机制要求每个计算步骤都使用较大的时间窗口，因此与循环神经网络（RNN）相比，不太适合在线信号处理。本文没有采用自注意力机制，而是使用滑动窗口注意力机制。我们证明，对于输入与目标之间具有有限范围依赖的连续信号，该机制效率更高；它还支持逐元素处理序列，因此适用于在线处理。我们在手指位置回归数据集（NinaproDB8）上测试了模型，利用测量自前臂皮肤的表面肌电（sEMG）信号来估计肌肉活动。就该数据集的准确率而言，我们的方法达到了新的最先进水平，同时每个推理步骤所需的时间窗口仅为 3.5 ms。此外，我们采用漏积分发放（LIF）单元提高网络的稀疏性。LIF 是一种仿生神经元模型，仅在越过阈值时才会稀疏地激活。由此，我们将突触运算数量减少了至多 $\times 5.3$ 倍，同时不损失准确率。我们的结果为实现准确、快速的 sEMG 信号在线处理，从而平稳控制假手带来了巨大希望，也为 Transformer 与脉冲神经网络（SNN）的协同集成迈出了一步，助力实现能效型时序信号处理。

1 引言

表面肌电图（表面肌电）是一种通过感测流经肌纤维膜的电流来测量肌肉活动的技术 [1]。由于表面肌电信号由中枢神经系统发出的电刺激触发，这种方法正日益受到关注，可用于人机交互 [1]。表面肌电测量只需将电极放置在前臂皮肤上，因此，这项技术在未来的非侵入式可穿戴假肢手控制系统中前景广阔 [1]。

预印本，正在评审中。

Transformer 是当前最先进的序列处理网络 [2, 3], ，能够高效处理表面肌电信号 [4]。然而，传统 Transformer 中使用的自注意力机制 [2]需要等待较长的时间窗口，从而造成延迟，妨碍连续信号的快速在线处理。此外，自注意力机制的内存占用和计算量会随序列长度呈二次方增长。

相比之下，循环神经网络（RNN）将时间概念融入其运行模式，因此适用于连续信号的在线处理。脉冲神经网络（SNN）[5, 6] 是一种受生物启发的循环神经网络。它们在低功耗应用中前景广阔，因为只有当神经元的膜电位（每个神经元的内部状态）达到阈值时，神经元才会传递信息，而且这些事件在时间上稀疏发生 [6]。许多研究致力于构建能够利用脉冲神经网络固有时间稀疏性的全新硬件 [7, 8, 9, 10]。

本文提出一种在线Transformer，采用线性化滑动窗口注意力机制 [11]。我们对这种注意力机制进行调整，使其沿时间向前推进并以串行方式处理，从而适用于连续信号的在线处理。因此，我们的在线Transformer会在每个词元生成时对其进行推理。为了利用过去输入中的信息，我们将信息存储在注意力机制的键和值中，并随着词元的生成动态更新这些内存。键和值中所存储序列的长度是一个可调节的超参数；通过调整它，我们可以改变注意力机制所利用信息的时间深度，以及计算复杂度和内存使用量。

我们使用非侵入式自适应手部假肢数据库8（NinaProDB8）数据集，通过表面肌电信号进行手指位置回归，以测试我们的模型。首先，我们证明，在线Transformer仅使用极短的3.5 ms时间窗口，就能让用户以高准确率处理表面肌电信号，从而实现时间粒度非常精细的假手控制。其次，我们证明，选择注意力的时间深度能够改善信号处理结果，使我们的模型优于基于自注意力的Transformer以及此前最先进的模型。最后，我们展示了定制的在线注意力机制如何使Transformer架构中的脉冲神经网络提高网络稀疏性，进而将所需的突触运算次数减少 $\times 5.3$ 倍，同时不损失准确率。

2 相关工作

用于表面肌电图处理的深度学习 尽管表面肌电信号与肌肉活动相关，但二者之间的关系尚不明确；由于存在电噪声（例如干扰、地噪声、电极间串扰）、受试者间变异（例如前臂周长和肌肉特征不同）以及受试者内变异（例如电极位置或皮肤电导率逐日变化），处理表面肌电信号仍然极具挑战性 [12]。

深度学习方法能够利用大型数据集，在存在噪声或变化的情况下提取更相关的特征 [13]。因此，它们的表现可以优于支持向量机（SVM）等传统机器学习技术 [14]。此外，深度网络能够处理原始表面肌电信号，而传统网络则需要先进行预处理，例如主成分分析 [15], 、线性判别分析（LDA） [15], 、傅里叶变换 [16],等。

深度学习已应用于基于时间卷积[17, 18, 19] 和循环神经网络（RNN） [20, 21, 22, 23]的表面肌电信号处理。尽管在内存容量有限、功耗较低的边缘设备上进行计算，是部署自主可穿戴假肢手控制系统的必要条件，但大多数深度学习技术的计算开销都很大。

Mukhopadhyay 等人 [24] 证明，可以利用脉冲神经网络固有的稀疏性，大幅降低表面肌电信号处理的计算开销。Burrello 等人 [4] 则证明，Transformer 网络能够以有限的内存使用量和较少的乘累加（MAC）运算处理表面肌电信号。

Transformer 与循环神经网络不同，Transformer 在学习时间特征时不会遭遇梯度消失问题 [3], ；它们没有基于数据结构假设形成的归纳偏置，而且由于能够并行处理整个时间序列，因此可以在图形处理器上快速训练。Transformer 的核心是自注意力机制，这种操作可以让序列中的所有元素相互比较。对于自然语言

处理（自然语言处理），自注意力的优势在于，它允许一个词元与当前、过去和未来的词元进行比较 [2]。然而，根据应用场景的不同，传统自注意力并非总是最佳选择。研究表明，在自然语言处理 [11] 和机器视觉 [25]中，对长序列使用局部注意力和滑动窗口注意力可以取得更好的结果。

采用脉冲神经网络的Transformer。脉冲神经网络模仿生物神经网络，非常适用于低功耗应用，因为只有当神经元的膜电位（每个神经元的内部状态）达到阈值时，神经元才会传递信息，而这些事件在时间上呈稀疏分布 [6]。将脉冲神经网络集成到Transformer架构中颇具挑战性，也并非直观可行。与循环神经网络一样，脉冲神经网络具有时间动态。因此，序列中的每个元素都必须依次输入循环神经网络或脉冲神经网络。相比之下，自注意力机制会并行比较序列中的所有不同元素，因此Transformer必须等待整个序列处理完毕后才能进行计算。例如， [4]中用于表面肌电分类的Transformer使用了150 ms的时间窗口。若简单地堆叠传统自注意力层和循环层，由于需要交替等待时间窗口和按顺序处理序列，将导致不理想的延迟。

Yao等人 [26] 采用一种注意力机制来选择事件帧的重要性，然后使用脉冲神经网络处理这些事件。Sabater等人 [27] 表明，与卷积神经网络相比，Transformer能够更高效、更准确地处理基于事件的视觉传感器数据。Zhou等人 [28] 采用二值化自注意力，将稀疏性融入Transformer。Li等人 [29] 将脉冲神经网络用作Transformer的预处理步骤。Gehrig和Scaramuzza [30] 也证明，长短期记忆网络单元可以集成到Transformer架构中，但在这项工作中，注意力机制关注的是空间而非时间。最后，Zhu等人 [31] 将脉冲神经元集成到Transformer架构中，但采用了无法在线计算的自定义注意力机制。

本文介绍了一种 Transformer 模型，该模型能够在线进行时间注意力计算，并且其架构的每一层都兼容脉冲神经元。

3 方法

3.1 NinaproDB8：手指位置回归数据集

在本研究中，我们使用了非侵入式自适应手部假肢数据库 8（NinaProDB8） [12], ，这是一个公开的表面肌电数据库，旨在作为手指运动学位置估计的基准。许多应用于表面肌电的深度学习研究都聚焦于简单的功能性动作分类 [1, 4, 17, 18]。然而，手指位置的序列到序列回归可以支持更广泛的手势，也更容易与机器人手的感觉反馈结合，从而实现闭环精确控制 [33]。

该数据库的测量数据采集自 10 名健全受试者和 2 名右侧经桡骨截肢者。表面肌电信号是我们神经网络的输入（见图1 (a)），通过 16 个电极（Delsys Trigno IM 无线肌电系统）记录，这些电极安置在参与者的右前臂周围。手指位置通过数据手套 Cyberglove II 进行测量，该手套为 18 自由度（DoF）型号，可测量与图1 (b) 中各点对应的手指关节角度。表面肌电信号和数据手套信号均上采样至 2 kHz，并进行后同步。有关该数据集的详细信息，请参见 [12]。

为了忽略无关的自由度，并直接关注与假手控制相关的动作，Krasoulis 等人 [12] 已证明，我们可以使用简单的线性变换，将数据手套记录的 18 个自由度转换为 5 个驱动自由度（DoA）。该线性变换所用的矩阵可在 Krasoulis 等人 [12]的补充材料中找到。我们将 DoA 用作神经网络的目标值。
每位参与者均记录了三个数据集：前两个数据集（采集 1 和 2）包含每种动作的 10 次重复，第三个数据集（采集 3）仅包含 2 次重复。我们将采集 1 和 2 用作训练集，将采集 3 用作测试集。图1 (a) 和 (c) 展示了受试者 1 的测试集示例（目标值）。

为便于训练神经网络，我们对每组重复动作进行归一化：先减去表面肌电信号的均值，再除以其标准差。

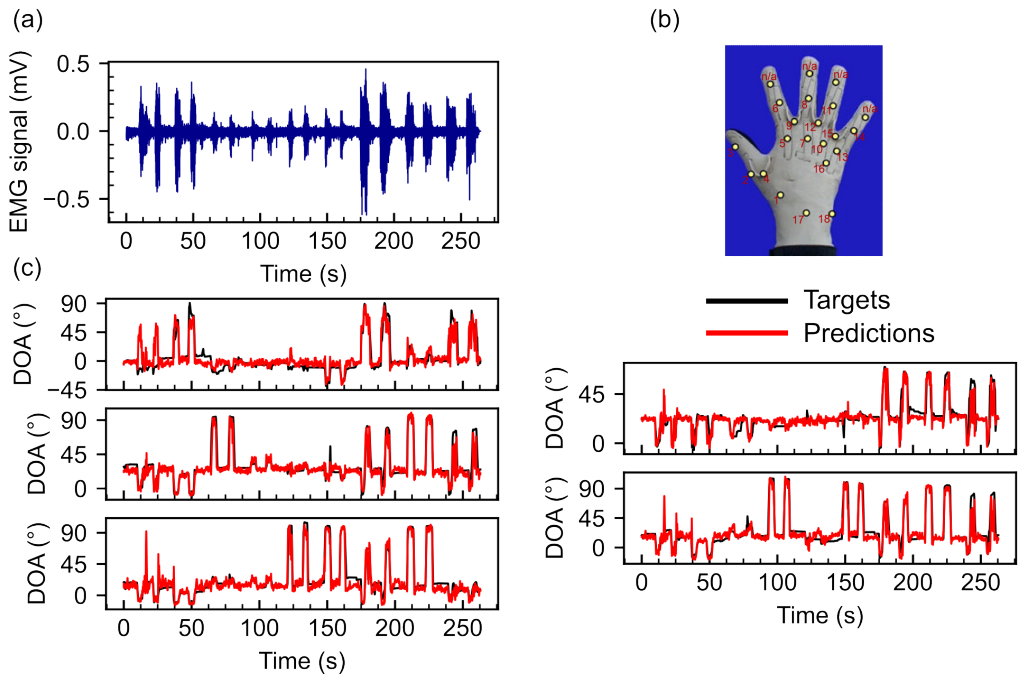

图 1: （a）使用 16 通道 DelsysTrigno IM 无线肌电系统采集的表面肌电信号（参见 Krasoulis 等人 [12]）。图中仅显示 16 个通道中的一个通道信号。（b）使用 Cyberglove II 采集真实手指关节角度 [32]。（c）真实手指关节角度及使用我们的在线 Transformer 模型进行的重建结果。

3.2 使用自定义注意力机制进行在线推理

在传统Transformer中， [2], 整个自注意力阶段都是并行计算的。自注意力层输入序列中的元素称为词元，自注意力的运算表示为

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{Q \circ K^T}{\sqrt{d}}\right) \circ V \tag{1}$$

其中， $\circ$ 表示点积运算符， $Q \in \mathbb{R}^{N\times d}$、 $K \in \mathbb{R}^{N\times d}$和 $V \in \mathbb{R}^{N\times d}$ 分别称为查询、键和值，它们是同一词元序列的三种不同投影：

$$Q = W_Q x \tag{2a}$$
$$K = W_K x \tag{2b}$$
$$V = W_V x. \tag{2c}$$

注意力维度 d 是每个词元投影的大小， N 是序列长度。 $W_Q \in \mathbb{R}^{d\times D}$、 $W_K \in \mathbb{R}^{d\times D}$和 $W_V \in \mathbb{R}^{d\times D}$ 是可学习权重矩阵，其中 D 是嵌入维度， $x \in \mathbb{R}^{N\times D}$ 是注意力机制的输入。

对于连续信号（例如生物医学信号），可以将输入信号划分为有限个时间窗口，并等待每个时间窗口结束后再进行推理（如 Burrello 等人 [4]所述）。然而，这种方法会因等待时间窗口结束而产生延迟。我们的在线Transformer采用了一种自定义注意力机制，可以对序列中的每个元素进行在线计算，而不会产生延迟。

为避免等待未来的词元，时间步 t_0 的词元不会与时间步 $t > t_0$之后的未来词元进行比较。来自先前词元的信息存储在键和值 $K \in \mathbb{R}^{M\times d}$ 与 $V \in \mathbb{R}^{M\times d}$中。与自注意力不同，这里的 K 和 V 的大小不取决于完整的序列长度，而仅取决于 M，即我们选择存储的过去时间步数。

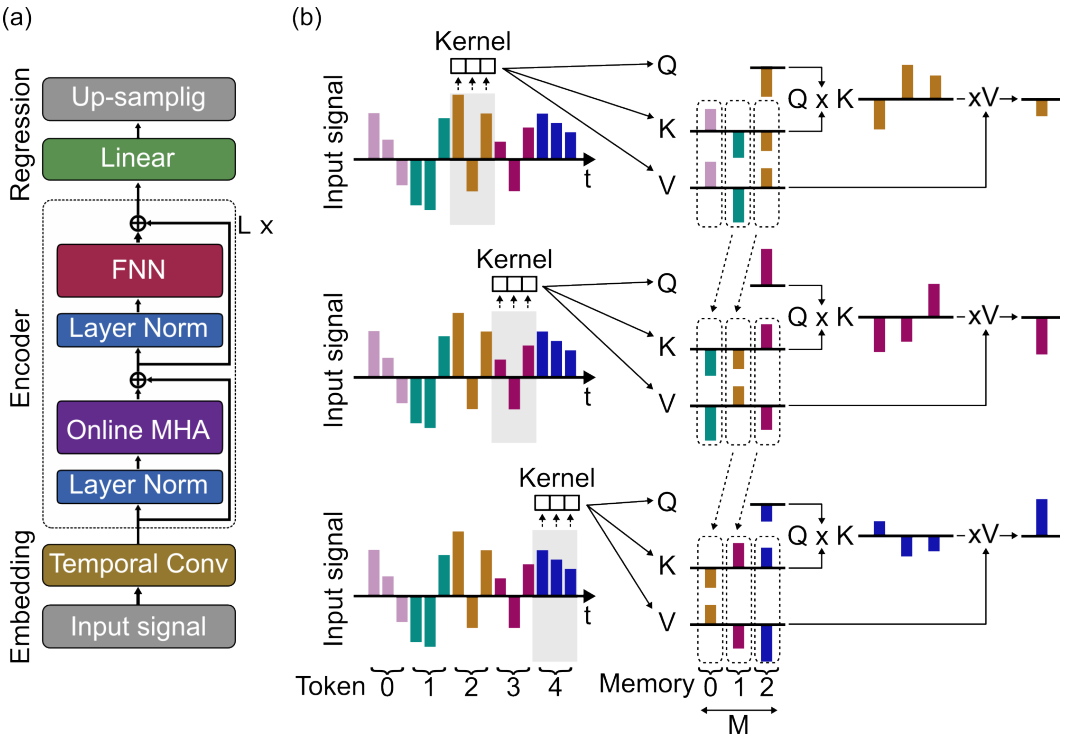

图 2：(a) 在线 Transformer 神经网络架构。(b) 在线注意力示意图：不同的词元由时间卷积生成（本例中的卷积核大小为 3，步幅为 2）。这些词元经过线性投影，分别生成查询、键和值（Q、 K、 V）。Q 仅与当前词元匹配，而 K 和 V 则存储多个先前的词元。本例中，该内存的长度 M 为 3。在每个时间步， K 和 V 会丢弃最旧词元的投影，并存储新词元的投影。在线注意力机制的数学运算将在第 3.2 节中介绍。

K 和 V 最初均置为零。随后，使用公式 2b 和 2c 中的投影，逐个词元迭代替换元素 $K_i \in \mathbb{R}^d$ 和 $V_i \in \mathbb{R}^d$。在每个时间步，还会根据公式 2a 计算单个查询 $Q_t \in \mathbb{R}^d$，并将注意力计算为

$$\text{Attention}_t = \text{softmax}\left(\frac{Q_t \circ K}{\sqrt{d}}\right) \circ V \tag{3}$$

Softmax 在记忆长度维度 M 上计算。由于每个时间步都会更新 K 和 V 中的一个不同元素，且 K 和 V 的长度为 M，因此它们的所有元素都会以 $\frac{1}{M}$ 的频率更新。上述过程总结为算法 1。

与传统自注意力（$O\left(N^2\right)$）对序列长度呈二次依赖不同，滑动窗口注意力机制的计算复杂度为线性（$O\left(MN\right)$）。

公式 3 表明，我们的注意力机制可以按时间步计算，而不必等待较长时间窗口结束后再并行计算注意力。接下来，我们将展示该注意力机制如何融入完整的神经网络。

3.3 神经网络架构

如图 2 (a) 所示，我们的神经网络由三个模块组成：将原始肌电信号转换为词元序列的嵌入块、利用注意力查找序列元素间相关性的编码器块，以及将输出转换为五个驱动自由度的回归块。

算法 1使用在线注意力进行推理

$K = \mathbf{0}$ ▷ $K \in \mathbb{R}^{M \times d}$
$V = \mathbf{0}$ ▷ $V \in \mathbb{R}^{M \times d}$
$t = 0$
$i = 0$
重复
  $Q_t \leftarrow W_Q x_t$ ▷ $Q_t \in \mathbb{R}^d$
  $K_i \leftarrow W_K x_t$ ▷ $K_i \in \mathbb{R}^d$
  $V_i \leftarrow W_V x_t$ ▷ $V_i \in \mathbb{R}^d$
  注意力$_t \leftarrow \text{Softmax}\left(\frac{Q_t \circ K}{\sqrt{d}}\right) \circ V$
  $t \leftarrow t + 1$ ▷ 新时间步
  $i \leftarrow i + 1$
  $i \leftarrow i$ (模M)
直到序列结束

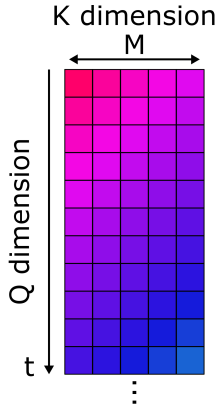

图3：滑动窗口注意力。相同的 K 值用相同的颜色表示。

嵌入由一个时间卷积层构成。卷积层的输入时间窗口之间存在重叠，这意味着与线性层不同，卷积层具有内在的顺序，因此不需要位置嵌入 [2]。卷积层具有与 16 个电极匹配的 $C = 16$输入通道，以及 $D = 64$输出通道。我们测试了不同大小的卷积核，以改变与单个词元对应的输入时间窗口长度。我们选择使时间窗口之间重叠两个时间步，因此卷积层的步幅为 $s = k - 2$，其中 k 是卷积核大小。采用填充 $p = 1$ 后，生成的词元数量取决于步幅，公式如下：

$$N = \lfloor \frac{N_{\text{samples}}}{s} \rfloor \tag{4}$$

其中， N_{samples} 是输入信号中已处理样本的数量。
编码器的描述如下：

$$\begin{aligned} f(x) &= x + \text{MHA}(\text{LN}(x)) \\ z(f(x)) &= f(x) + \text{FNN}(\text{LN}(f(x))) \end{aligned} \tag{5}$$

其中，LN 是层归一化层。 MHA 是多头注意力层，包含 $h = 8$ 个注意力头，这些注意力头依据公式 3 并行计算，注意力维度为 $d = 32$。完成注意力计算后，将h 个注意力头拼接起来，并投影至维度 $D = 64$。 FNN是一个前馈神经网络，包含一个由 128 个 GeLU 单元组成的隐藏层 [34] ，以及一个丢弃概率为 0.2 的丢弃层。前馈神经网络的线性投影逐词元应用，因此可以在线计算。整个编码器块可以重复并堆叠 L 次，但在这里我们选择保留 $L = 1$。我们的神经网络主干受到 Burrello 等人 [4]的启发。
最后，回归块由一个线性层和一个上采样层组成。线性层将每个词元从维度$D = 64$ 投影至维度 5（即我们进行回归的驱动自由度数量）；上采样层则复制每个词元的输出，生成与目标信号样本数相同的样本（目标信号示例见图1 (c)）。上采样因子等于我们在卷积嵌入层中使用的步幅（见公式 4）。

3.4 通过二值化和脉冲神经元提高网络稀疏性

为了减少所需运算的数量，我们通过二值化和漏积分发放（LIF）单元 [5, 6] units 来提高网络稀疏性。我们测试了两种稀疏模型。第一种模型将卷积嵌入的输出二值化，将投影 Q、 K 和 V 二值化，并以脉冲神经网络（SNN）取代前馈神经网络（FNN）：第一层包含 128 个 LIF 单元，第二层包含 $D = 64$ 个 LIF 单元。第二种稀疏模型与第一种类似，但它不对 Q、 K 和 V 进行二值化，而是将式 2a、2b 和 2c 中的每个投影替换为一个包含 $d = 32$ 个 LIF 单元的脉冲层，这为模型增添了额外的动态特性。

Bin二值化通过应用赫维赛德函数实现。LIF 单元的动态特性定义如下

$$U_t = \alpha\left(1 - S_{t-1}\right) U_{t-1} + (1-\alpha)\, I_{t-1} \tag{6a}$$

$$I_t = \beta I_{t-1} + (1-\beta)\, W x_t \tag{6b}$$

$$S_t = H\left(U_{t-1} - \Theta\right) \tag{6c}$$

其中， t是词元的索引， U 是膜电位， I 是突触电流，S 是脉冲响应，H 是赫维赛德函数， $\alpha = 0.95$、 $\beta = 0.9$和 $\boldsymbol{\Theta} = 1$。Q、 K和 V 投影的输出是脉冲响应 S_t （见公式 6c）。替代 FNN 的 SNN 第一层的输出是脉冲响应 S_t，第二层的输出是膜电位 U_t （见公式 6a ）。

由于赫维赛德函数不可微，在训练过程中，不同赫维赛德函数（用于二值化和 LIF 单元）的梯度由 SuperSpike 替代梯度 [5, 35]取代。为保留嵌入与编码器块之间的稀疏性，当嵌入进行二值化时，我们移除嵌入之前的层归一化层。此外，我们还移除了两个稀疏模型中的丢弃层。注意力机制的 Softmax 仅在非零元素上计算。

3.5 训练

为了加快训练，注意力块采用并行计算。对于包含 N 个词元的整个时间窗口，计算投影 $Q \in \mathbb{R}^{N\times d}$、 $K \in \mathbb{R}^{N\times d}$ 和 $V \in \mathbb{R}^{N\times d}$。随后，将键和值展开为大小为 M、步幅为 1 的滑动窗口，方法与卷积类似（滑动窗口注意力示例见图 3）。因此，查询与键的乘积计算如下:

$$\begin{bmatrix} Q_0K_{1-M} & \cdots & Q_0K_{-2} & Q_0K_{-1} & Q_0K_0 \\ Q_1K_{2-M} & \cdots & Q_1K_{-1} & Q_1K_0 & Q_1K_1 \\ Q_2K_{3-M} & \cdots & Q_2K_0 & Q_2K_1 & Q_2K_2 \\ \vdots & \iddots & \vdots & \vdots & \vdots \\ Q_NK_{N-M} & \cdots & Q_{N-1}K_{N-3} & Q_{N-1}K_{N-2} & Q_{N-1}K_{N-1} \end{bmatrix}. \tag{7}$$

由于键 $K_{i<0}$ 是禁止值，我们按照 Vaswani 等人 [2], 的方法，将其替换为 $-\infty$ 进行屏蔽，使其不参与 Softmax 计算（见公式 3）。值 $V_{i<0}$ 则直接置为零。

为了改进训练，我们设计了一种简单的数据增强方案：首先，将训练集中的信号切分为长度为 $N_{\text{samples}} = 2000$ 个采样点的时间窗口（由于采样率为 2 kHz，这相当于 1 秒）。然后，将每个时间窗口复制 64 次。进行数据增强时，在 0 到 2000 之间的均匀分布中随机抽取一个数，并据此偏移每个复制所得时间窗口的起始位置。最后，将得到的时间窗口打乱，以创建训练数据集。

我们使用 Adam 优化器 [36], 和学习率 10^{-3}、批量大小 64，为每位受试者训练每个网络 10 个 epoch。由于我们希望最小化的指标是 5 个自由度（DoA）上的平均绝对误差（MAE），因此使用了 L1 损失函数。

对于稀疏模型，我们在全局损失中添加了稀疏性损失函数项 [37] ，以提高嵌入、查询、键和值的稀疏性，使总损失为:

$$\mathcal{L}\left(y, \hat{y}\right) = \| \, y_{i,j} - \hat{y}_{i,j} \, \|_1 - \frac{1}{2}\lambda\left(\| \, x \, \|_2 + \| \, \mathrm{Concat}\left(Q, K, V\right) \, \|_2\right) \tag{8}$$

其中 y 为网络输出，$\hat{}y$ 为目标值， x 为嵌入，且 $\lambda = 1$。

在本研究中，我们仅针对每位受试者分别独立训练和测试数据集。为提高未来研究中的准确率和可重复性，也可以采用迁移学习：网络可以先从多位受试者的数据中学习，再针对新的受试者进行微调和测试，如 [38] 所示。

4 结果

图1 (c) 展示了一个稀疏在线Transformer的回归结果示例，其嵌入卷积核大小为 $k = 7$，记忆长度为 $M = 150$。我们首先考察 k

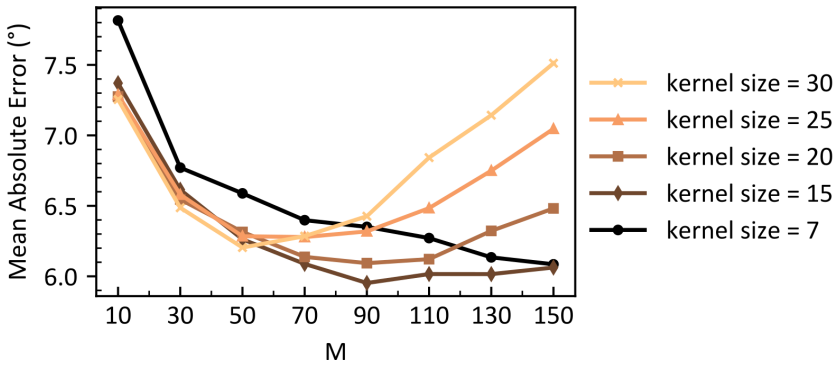

图4：在12名受试者中取平均的平均绝对误差与存储词元数量 M 的关系；卷积核嵌入尺寸分别为7、15、20、25和30（曲线颜色由黑色渐变至浅黄色）。

和 M 对最终准确率的影响。在这项研究中，我们使用不具稀疏性的网络。由于 $s = k - 2$，我们同时改变 s 和 k，因此对于给定的时间窗口生成的词元数量为 N（见式4）。记忆长度 M 定义了注意力机制中使用多少个过去的词元。因此，用于在 K 和 V 中存储信息的信号时间长度为：

$$\tau_{\text{memory}} = \frac{M \times s}{\text{SamplingRate}}. \tag{9}$$

在图4中，我们绘制了不同驱动自由度下的平均绝对误差（平均绝对误差），其中 M 在10至150之间以20为间隔变化，并采用五种不同的 $k = 7$ 值：15、20、25和30（分别对应 $s = 5$、13、18、23和28）。在改变 M 时，我们发现对于 $k = 15$、20、25和30，平均绝对误差先达到最小值，随后增大。这表明每种卷积核大小都存在一个最佳记忆长度，并且该最佳值会随卷积核大小增大而减小，因此也会随步幅增大而减小。根据公式9，这一结果表明注意力机制中所使用的信息长度 τ_{memory} 存在最佳值；超过该值后，增加存储的信息并不会提高准确率。接下来，我们比较不同模型，每次分别采用卷积核大小 $k = 7$ 和记忆长度 $M = 150$。如图4所示，对于每个词元，这些参数在采用最短时间窗口时可获得最佳准确率。在本研究中，我们还测量了10°准确率和15°准确率，分别表示平均绝对误差小于10度和15度的时间样本所占比例 [21]。这些补充指标对于衡量预测在误差范围内的准确率至关重要。不同结果见表4。每项指标的均值和标准差均根据NinaproDB8 数据集中12名受试者的数据计算得出。

为了考察自定义滑动窗口注意力机制对在线 Transformer 的影响，我们将其与采用自注意力机制的传统Transformer进行比较。在三项指标上，我们的在线 Transformer 均优于采用传统自注意力的 Transformer（见表4）。这些结果进一步揭示了选择相关信息的重要性；对于表面肌电信号处理而言，利用过去的局部信息可能比利用来自过去和未来的全局信息更为重要。

我们的两种稀疏模型取得了与非稀疏在线 Transformer 相近的准确率（因此也优于对应的传统 Transformer），并且与非稀疏在线 Transformer 相比，所需乘累加运算（乘累加运算）的数量分别减少至原来的 $3.8\times$ 和 $5.3\times$倍（这些计算未包含激活函数运算）。计算所需运算次数的方法见附录。

此外，我们发现，三个在线 Transformer 模型的平均绝对误差至少比长短期记忆网络低 0.88°，比时序卷积至少低 0.76°（此前 NinaproDB8 数据集上的最先进水平）。为比较不同方法的推理速度，我们将计算所需的最短时间定义为每个推理步骤所用时间窗口的长度。对于我们的在线 Transformer，该长度为 $\tau_{\min} = \frac{k}{\text{采样率}}$，其中 k 为嵌入卷积核大小。由于 $k = 7$ 且采样率为 2 kHz，我们的网络能够以最低延迟进行计算，延迟为

表 1：使用不同模型在 Ninapro DB8 数据库上进行回归的结果

模型	**平均绝对误差 (°)↓**	**10°准确率↑**	**15°准确率↑**	$\tau_{\mathbf{min}}$ ↓	**百万乘加运算次数↓ /推理**
SVM* [19]	7.28	0.79	0.88	60 毫秒	–
多层感知机* [19]	7.14	0.80	0.89	60 毫秒	–
TempConv [19]	6.89	0.81	0.90	128 毫秒	3.2
长短期记忆网络 [21]	7.04	–	–	10 毫秒	–
Transformer†	6.62±1.52	0.87±0.07	0.92±0.05	2 s	3,769
在线Transformer†	**6.10±1.50**	**0.86±0.07**	**0.94±0.05**	**3.5 毫秒**	5.3
在线Transformer与二值嵌入和QKV, 脉冲前馈神经网络†	**6.08±1.27**	**0.87±0.06**	**0.94±0.04**	**3.5 毫秒**	**1.4**
采用二值嵌入, 脉冲式 QKV 和前馈神经网络†	**6.16±1.39**	**0.87±0.07**	**0.94±0.04**	**3.5 毫秒**	**1.0**

* 采用先前的特征提取。
† 本研究。

3.5 ms，比以往任何方法都短，尤其比此前在 Ninapro DB8 数据集上达到最先进水平的时序卷积网络 [19] 短了超过 30× 。

5 结论

在本研究中，我们开发了一种在线 Transformer 模型，利用滑动窗口注意力逐个处理词元。我们证明，滑动窗口的局部性使其比自注意力更高效。所提出的方法使得使用极短的时间窗口（3.5 毫秒）进行表面肌电信号处理成为可能，并在用于假手控制的 NinaproDB8 数据集上取得了新的最先进水平。借助滑动窗口注意力，我们的模型还解决了在 Transformer 中整合脉冲神经网络时间动态的问题。我们结合二值化与脉冲神经网络来提高网络稀疏性，从而将所需运算次数最多减少至原来的5.3×倍。总之，本研究迈出了实现精确、平稳、低功耗人机交互的一步，并为未来的神经形态 Transformer 模型带来了巨大希望。

参考文献

[1] Mingde Zheng、Michael S. Crouch 和 Michael S. Eggleston。表面肌电图作为自然的人机接口：综述。22(10):9198–9214。ISSN 1558-1748。doi: 10.1109/JSEN.2022.3165988。期刊名称：IEEE Sensors Journal。[2] Ashish Vaswani、Noam Shazeer、Niki Parmar、Jakob Uszkoreit、Llion Jones、Aidan N Gomez、Łukasz Kaiser 和 Illia Polosukhin。Attention is all you need。载于神经信息处理系统进展，第30卷。Curran Associates, Inc. 网址：https://papers.nips. cc/paper/2017/hash/3f5ee243547dee91fbd053c1c4a845aa-Abstract.html。[3] Tianyang Lin、Yuxin Wang、Xiangyang Liu 和 Xipeng Qiu。Transformer综述。3:111–132。ISSN 2666-6510。doi: 10.1016/j.aiopen.2022.10.001。网址：https: //www.sciencedirect.com/science/article/pii/S2666651022000146。[4] Alessio Burrello、Francesco Bianco Morghet、Moritz Scherer、Simone Benatti、Luca Benini、Enrico Macii、Massimo Poncino 和 Daniele Jahier Pagliari。Bioformers：用于超低功耗基于表面肌电的手势识别的嵌入式Transformer。载于*2022*年欧洲设计、自动化与测试会议暨展览会（*DATE*），第1443–1448页。doi: 10.23919/DATE54114.2022.9774639。ISSN：1558-1101。

Python 代码可在 https://github ·com/NathanLeroux-git/在线脉冲神经元变换器

[5] Friedemann Zenke 和 Surya Ganguli。SuperSpike：多层脉冲神经网络中的监督学习。30(6):1514–1541。ISSN 0899-7667。doi: 10.1162/neco_a_01086。网址：https://doi.org/10.1162/neco_a_01086。[6] Amirhossein Tavanaei、Masoud Ghodrati、Saeed Reza Kheradpisheh、Timothée Masquelier 和 Anthony Maida。脉冲神经网络中的深度学习。111:47–63。ISSN 0893-6080。doi: 10.1016/j.neunet.2018.12.002。网址：https://www.sciencedirect.com/science/ article/pii/S0893608018303332。[7] Paul A. Merolla、John V. Arthur、Rodrigo Alvarez-Icaza、Andrew S. Cassidy、Jun Sawada、Filipp Akopyan、Bryan L. Jackson、Nabil Imam、Chen Guo、Yutaka Nakamura、Bernard Brezzo、Ivan Vo、Steven K. Esser、Rathinakumar Appuswamy、Brian Taba、Arnon Amir、Myron D. Flickner、William P. Risk、Rajit Manohar 和 Dharmendra S. Modha。包含可扩展通信网络和接口的百万脉冲神经元集成电路。345(6197): 668–673。ISSN 0036-8075、1095-9203。doi: 10.1126/science.1254642。网址：https://science.sciencemag.org/content/345/6197/668。期号：6197。出版方：美国科学促进会。栏目：报告。[8] Chen Liu、Guillaume Bellec、Bernhard Vogginger、David Kappel、Johannes Partzsch、Felix Neumärker、Sebastian Höppner、Wolfgang Maass、Steve B. Furber、Robert Legenstein 和 Christian G. Mayr。在SpiNNaker 2原型机上实现内存高效深度学习。12。ISSN 1662- 453X。网址：https://www.frontiersin.org/articles/10.3389/fnins.2018.00840。[9] Garrick Orchard、E. Paxon Frady、Daniel Ben Dayan Rubin、Sophia Sanborn、Sumit Bam Shrestha、Friedrich T. Sommer 和 Mike Davies。采用Loihi 2实现高效的神经形态信号处理。载于*2021*年*IEEE*信号处理系统研讨会（*SiPS*），第254–259页。doi: 10.1109/SiPS52927.2021.00053。ISSN：2374-7390。

[10] Christian Pehle、Sebastian Billaudelle、Benjamin Cramer、Jakob Kaiser、Korbinian Schreiber、Yannik Stradmann、Johannes Weis、Aron Leibfried、Eric Müller 和 Johannes Schemmel。《采用混合可塑性的 BrainScaleS-2 加速神经形态系统》。16:795876。ISSN 1662-4548。doi: 10.3389/fnins.2022.795876。URL https://www.ncbi.nlm.nih.gov/pmc/ articles/PMC8907969/。

[11] Iz Beltagy、Matthew E. Peters 和 Arman Cohan。Longformer：长文档 Transformer。网址：http://arxiv.org/abs/2004.05150。[12] Agamemnon Krasoulis、Sethu Vijayakumar 和 Kianoush Nazarpour。用户练习对使用直观肌电解码器进行假肢手指控制的影响。13。ISSN 1662-453X。网址：https://www.frontiersin.org/articles/10.3389/fnins.2019.00891。[13] Andrés Jaramillo-Yánez、Marco E. Benalcázar 和 Elisa Mena-Maldonado。使用表面肌电图和机器学习进行实时手势识别：系统性文献综述。20(9):2467。ISSN 1424-8220。doi：10.3390/s20092467。网址：https://www.mdpi.com/1424-8220/20/9/2467。期号：9。出版方：多学科数字出版研究所。[14] Bojan Milosevic、Elisabetta Farella 和 Simone Benatti。探索肌电手势识别中的手臂姿势与时间变异性。发表于*2018*年第七届*IEEE*生物医学机器人与生物机电国际会议（*Biorob*），第1032–1037页。doi：10.1109/BIOROB.2018.8487838。ISSN：2155-1782。[15] Pornchai Phukpattaranont、Sirinee Thongpanja、Khairul Anam、Adel Al-Jumaily 和 Chusak Limsakul。使用表面肌电信号进行手指运动识别的特征提取技术与分类器评估。56(12):2259–2271。ISSN 1741-0444。doi：10.1007/s11517-018-1857-5。网址：https://doi.org/10.1007/s11517-018-1857-5。[16] Zahra Taghizadeh、Saeid Rashidi 和 Ahmad Shalbaf。基于从表面肌电信号中提取的分数阶傅里叶变换系数进行手指动作分类。68:102573。ISSN 1746-8094。doi：10.1016/j.bspc.2021.102573。网址：https://www.sciencedirect.com/science/article/pii/S1746809421001701。

[17] Panagiotis Tsinganos、Bruno Cornelis、Jan Cornelis、Bart Jansen 和 Athanassios Skodras。基于表面肌电信号和时间卷积网络的改进手势识别。发表于 *ICASSP 2019 - 2019*年*IEEE*声学、语音与信号处理国际会议（*ICASSP*），第1169–1173页。doi：10.1109/ICASSP.2019.8683239。ISSN：2379-190X。[18] Marcello Zanghieri、Simone Benatti、Alessio Burrello、Victor Kartsch、Francesco Conti 和 Luca Benini。在多核物联网处理器上使用时间卷积网络的鲁棒实时嵌入式肌电识别框架。14(2):244–256。ISSN 1940-9990。doi：10.1109/TBCAS.2019.2959160。会议名称：IEEE生物医学电路与系统汇刊。[19] Marcello Zanghieri、Simone Benatti、Alessio Burrello、Victor Javier Kartsch Morinigo、Roberto Meattini、Gianluca Palli、Claudio Melchiorri 和 Luca Benini。使用低功耗边缘微控制器上的时间卷积网络进行基于表面肌电的手部运动学回归。发表于*2021*年*IEEE*全层智能系统国际会议（*COINS*），第1–6页。doi：10.1109/COINS51742.2021.9524188。[20] Khairul Anam、Cries Avian、Dwiretno Istiyadi Swasono、Aris Zainul Muttaqin 和 Harun Ismail。使用长短期记忆网络，基于肌电信号估计手指关节运动。发表于*2020*年计算机工程、网络与智能多媒体国际会议（*CENIM*），第86–90页。doi：10.1109/CENIM51130.2020.9298023。[21] Philipp Koch、Mark Dreier、Anna Larsen、Tim J. Parbs、Marco Maass、Huy Phan 和 Alfred Mertins。使用循环神经网络从表面肌电数据回归手部动作。发表于*2020*年第*42*届*IEEE*医学与生物学工程学会年度国际会议（*EMBC*），第3783–3787页。doi：10.1109/EMBC44109.2020.9176278。ISSN：2694- 0604。[22] Zamroni Ilyas、Khairul Anam、Widjonarko、Cries Avian、Aris Zainul Muttaqin 和 Mocham-mad Edoward Ramadhan。评估门控循环单元，以使用表面肌电信号估计手指关节角度。发表于*2022*年第*9*届电气工程、计算机科学与信息学国际会议（*EECSI*），第25–28页。doi：10.23919/EECSI56542.2022.9946461。[23] Jun Li、Lixin Wei、Yintang Wen、Xiaoguang Liu 和 Hongrui Wang。基于特征融合的表面肌电连续手部动作识别方法。ISSN 0178-2789，1432- 2315。doi：10.1007/s00371-022-02465-7。网址：https://link.springer.com/10.1007/ s00371-022-02465-7。[24] Anand Kumar Mukhopadhyay、Indrajit Chakrabarti 和 Mrigank Sharad。使用人工脉冲神经网络模型，通过表面肌电信号对手部动作进行分类。发表于*2018 IEEE SENSORS*，第1–4页。doi：10.1109/ICSENS.2018.8589757。ISSN：2168-9229。[25] Ali Hassani 和 Humphrey Shi。扩张邻域注意力变换器。网址：http: //arxiv.org/abs/2209.15001。[26] Man Yao、Huanhuan Gao、Guangshe Zhao、Dingheng Wang、Yihan Lin、Zhaoxu Yang 和 Guoqi Li。用于事件流分类的时间维注意力脉冲神经网络。第10221–10230页。网址：https://openaccess.thecvf.com/content/ ICCV2021/html/Yao_时间维_注意力_脉冲_神经_网络_用于_事件_流_分类_ICCV_2021_论文.html。[27] Alberto Sabater、Luis Montesano 和 Ana C. Murillo。事件变换器：一种面向稀疏数据、用于高效事件数据处理的解决方案。发表于*2022*年*IEEE/CVF*计算机视觉与模式识别会议研讨会（*CVPRW*），第2676–2685页。IEEE。ISBN 978-1-66548- 739-9。doi：10.1109/CVPRW56347.2022.00301。网址：https://ieeexplore.ieee.org/document/9857382/。[28] Zhaokun Zhou、Yuesheng Zhu、Chao He、Yaowei Wang、Shuicheng Yan、Yonghong Tian 和 Li Yuan。Spikformer：当脉冲神经网络遇上 Transformer。网址：http: //arxiv.org/abs/2209.15425。

[29] Yudong Li、Yunlin Lei 和 Xu Yang。Spikeformer：一种用于训练高性能、低延迟脉冲神经网络的新型架构。URL http://arxiv.org/abs/2211. 10686。[30] Mathias Gehrig 和 Davide Scaramuzza。用于事件相机目标检测的循环视觉 Transformer。URL http://arxiv.org/abs/2212.05598。[31] Rui-Jie Zhu、Qihang Zhao 和 Jason K. Eshraghian。SpikeGPT：采用脉冲神经网络的生成式预训练语言模型。URL http://arxiv.org/abs/2302.13939。[32] Stefano Pizzolato、Luca Tagliapietra、Matteo Cognolato、Monica Reggiani、Henning Müller 和 Manfredo Atzori。六种肌电图采集方案在手部运动分类任务中的比较。12(10):e0186132。ISSN 1932-6203。doi: 10.1371/journal.pone.0186132。URL https://dx.plos.org/10.1371/journal.pone.0186132。[33] Marko Markovic、Meike A. Schweisfurth、Leonard F. Engels、Dario Farina 和 Strahinja Dosen。肌电控制即闭环控制：偶然反馈足以在日常抓握中调节假肢作用力。15(1):81。ISSN 1743-0003。doi: 10.1186/s12984-018-0422-7。URL https://doi.org/10.1186/s12984-018-0422-7。[34] Dan Hendrycks 和 Kevin Gimpel。高斯误差线性单元（GELU）。URL http://arxiv.org/abs/1606.08415。[35] Emre O. Neftci、Hesham Mostafa 和 Friedemann Zenke。脉冲神经网络中的替代梯度学习：将基于梯度的优化能力引入脉冲神经网络。36(6):51–63。ISSN 1558-0792。doi: 10.1109/MSP.2019.2931595。编号：6。刊物名称：IEEE 信号处理杂志。[36] Diederik P. Kingma 和 Jimmy Ba。Adam：一种随机优化方法。URL http://arxiv.org/abs/1412.6980。[37] Yulong Yan、Haoming Chu、Yi Jin、Yuxiang Huan、Zhuo Zou 和 Lirong Zheng。通过稀疏正则化进行脉冲神经网络学习的反向传播。16。ISSN 1662-453X。URL https://www.frontiersin.org/articles/10.3389/fnins.2022.760298。[38] Stephan Johann Lehmler、Muhammad Saif-ur Rehman、Tobias Glasmachers 和 Ioannis Iossifidis。与受试者特定模型相比，深度迁移学习在表面肌电解码器中的应用。ISSN 1741-2560, 1741-2552。doi: 10.1088/1741-2552/ac9860。URL https://iopscience.iop.org/article/10.1088/1741-2552/ac9860。

A 附录

为了计算不同模型的突触 MAC 运算次数，我们使用了以下方程组来表示模型中的不同运算（不包括激活函数）：

$$\begin{aligned}
\text{\#嵌入乘累加} &= k \times C \times D \times 32 && \text{(10a)}\\
\#\text{QKVProjectionMACs} &= (1 - \text{EmbeddingSparsity}) \times 3 \times D \times d \times h \times 32 && \text{(10b)}\\
\text{\#QKProduct乘累加} &= \| QK \|_1 \times M \times h \times 32 && \text{(10c)}\\
\#\text{VProductMACs} &= (1 - \text{VSparsity}) \times d \times M \times h \times 32 && \text{(10d)}\\
\#\text{ConcatMACs} &= (1 - \text{AttentionSparsity}) \times d \times h \times D \times 32 && \text{(10e)}\\
\#\text{TotalAttentionMACs} &= \#\text{QKVProjectionMac} + \#\text{QKProductMacs} && \text{(10f)}\\
&\quad + \#\text{VProductMacs} + \#\text{ConcatMACs}\\
\text{\#FFL1乘累加} &= D \times H \times 32 && \text{(10g)}\\
\#\text{FFL2MACs} &= (1 - \text{FFL1Sparsity}) \times H \times D \times 32 && \text{(10h)}\\
\text{\#回归乘累加次数} &= D \times 5 \times 32 && \text{(10i)}\\
\#\text{TotalMACs} &= \#\text{EmbeddingMacs} + \#\text{TotalAttentionMACs} && \text{(10j)}\\
&\quad + \#\text{FFL1MACs} + \#\text{FFL2MACs} + \#\text{RegressionMACs}
\end{aligned}$$