

脑电研究：通过神经活动和视觉特征的多模态学习 解码大脑表征

方桂安^{*}, 王涛老师[†]

中山大学 智能科学与技术 20354027

【摘要】 课程设计的要求是针对认知科学基础学习的知识要点，从认知神经工程（生物学角度认知），到认知架构计算模型，再到认知自动化的工程应用，结合所讲飞行器制导控制系统引用的支持型和操作性人工认知单元设计思想方法，系统性地给出一种神经系统工程设计思考。作为智能工程学院的学生，我从本专业比较熟悉的机器学习、深度学习入手，研究学习了脑电相关的认知神经科学。

【关键词】 脑电信号，认知神经科学，深度学习

1 引言

在当下，人工系统几乎可以与人类的视觉能力相媲美，这主要归功于深度学习方面的最新进展。实际上，受到灵长类动物视觉系统的启发而产生的深层前馈和递归神经网络已显著提高了计算机视觉、自然语言处理、语音识别等的性能。除了在这类任务中获得表现的显著提高外，通过深度计算模型学习的表征似乎与大脑表征高度相关。例如，可以发现视觉通路中的大脑表征与深层神经网络 (DNNs) 中各层的层次结构之间的相关性。这些发现有利于认知神经科学家和人工智能研究人员参与多学科研究，目的是对人类大脑及其适应能力进行逆向工程。然而，这一多学科领域仍处于初级阶段。事实上，现有的计算神经模型虽然粗略地模拟了生物神经元的计算和连接，但它们往往忽略了前馈和反馈神经的相互作用。例如，人类的视觉识别似乎通过跨大脑皮质区域正向和反向加工的信息的多级聚合而得以缓解。

最近的一些方法受到神经科学中层次预测编码的启发，尝试通过提出具有前馈、反馈和循环连

接的递归神经网络来将这些额外的信息编码到计算模型中。这些模型在视觉分类任务中具有良好的表现，并表明更详细地了解人脑可能会让我们将知识转移到工程模型中，以创造更好的机器。显然，在将人类水平的分类性能转移到计算模型之前，首先需要更好地理解人类的视觉系统。为了实现这一点，可以将人类被试在执行特定任务时记录的神经活动数据与为完成相同任务而开发的计算模型相关联。通过研究学习到的计算表征以及它们如何随时间与神经活动相关，就有可能推断、分析并最终在机器中复制复杂的大脑加工。

查阅资料的过程中，我发现此类研究在近年来越来越多，许多研究者认为既然我们无法直接观察大脑，那么设计一个和大脑在任务表现和任务执行过程都相似的计算模型后，计算这个模型与人类表现的相似性，在极大相似的情况下，通过研究这个模型的具体运算过程可能有助于我们理解大脑，不论这种方法在假设上是否正确，但至少提出了一种靠近人脑的工程方法。

在本文中，首先提出了一个通过对大脑对自然图像的反应进行分类来学习神经表征的模型。然后，介绍了一种基于深度学习脑电和图像编码器的多模态方法，并在孪生网络下进行训练。利用几个对静止图像进行视觉分类的被试的脑电活动数据学习了一种联合的大脑-视觉特征，并发现了大脑表征和视觉特征之间的相似性。然后，该特征被

调研时间: 2022 年 7 月 10 日

报告时间: 2022 年 7 月 10 日

[†] 指导教师

^{*} 学号: 20354027

^{*} E-mail: fanggan@mail2.sysu.edu.cn

用于执行图像分类、显著性检测，并假设在人脑中生成的可能的表征用于视觉场景分析。

由此证明了：

a) 神经活动数据可用于为深度学习模型提供更丰富的监督标记，从而使视觉分类和显著性检测方法与人类神经数据保持一致；

b) 人工智能和认知神经科学的共同努力可能通过最大化深度模型与人类神经反应的相似性来揭示人类视觉感知中涉及的神经加工。事实上，本文提出了一种方法来提取视觉显著性（及其随时间的演变），以及定位产生这些信息的皮层区域；

c) 计算表征和大脑加工之间存在潜在的相似性，这为生物学和深度学习模型之间的一致性提供了有趣的见解。

2 相关工作

近年来，基于深度学习的脑电数据分类方法越来越流行。这些方法中的大多数都提出了定制的 AI 解决方案，用于对 BCI(脑机接口) 应用程序（例如，运动图像、语音想像、情感识别等）、临床应用程序（例如，癫痫检测和预测）或者监测认知功能（脑力负荷、参与、疲劳等）的数据进行分类。上述工作集中在少数几个类别的分类（从二进制分类到 10 个以下的类别），没有一个是理解人类视觉加工（在空间和时间上）为主要目标的。此外，大多数提出的模型已经应用于单一的脑机接口范式，具有非常可控的刺激。显而易见的担忧是，这些方法可能无法推广，并且随着刺激和任务的微小变化，它们的功能可能会崩溃。本文设计的模型试图解决这一泛化问题。该方法在 P300 视觉诱发电位、错误相关负反应 (ERN)、运动相关皮层电位 (MRCP) 和感觉运动节奏 (SMR) 等任务中表现良好。

本文提出的 EEG 分类方法旨在通过对神经反应的更一般的时空特征进行建模来支持认知神经科学的研究，以提高人类神经数据在时间和空间上的可解释性。

2.1 用于解码大脑表征的计算神经科学

破译大脑表征一直是这个时代的巨大挑战。认知神经科学在理解起源于初级视觉皮层 (V1) 的神经表征方面取得了很大进展。实际上，已知初级视觉皮质是在视网膜定位组织上的一系列定向边

缘和颜色检测器，其信号前馈到专注于更复杂的形状和特征维度的神经区域，这些神经区域在区域 V4 中较大的感受野上起作用，最后到达下颞叶 (IT) 皮层中的对象和类别表征。神经影像学方法，例如 fMRI，MEG 和 EEG，对于这些发现至关重要。然而，要重建完全代表我们视觉加工的人类水平的神经表征，需要同时精确地监测大脑中每个神经元的活动。尽管这些方法显然无法实现这一崇高目标，但它们包含足够的信息以准确地重建许多视觉体验。为此，大脑表征解码最近考察了神经活动数据和计算模型之间的相关性。然而，这些方法主要是在深度学习表征和神经成像数据之间进行简单的相关，并根据获得的结果得出关于大脑表征的结论，这过于简单化了。

对人的视觉系统的了解将来自训练自动模型以最大程度地增强大脑活动与激发刺激之间的信号相关性，而不是对大脑活动数据进行纯粹的分析。此外，虽然大多数方法试图使用来自高空间分辨率 fMRI 的脑图像来解码大脑表征，但采用脑电数据的方法，尽管空间分辨率较低，但具有较高的时间分辨率，这使得它更适合于解码快速的大脑加工，比如那些涉及视觉通路的加工。此外，与 fMRI 不同的是，脑电图是便携的、可移动的，甚至可以无线使用，这一特点将改善任何脑机接口。

2.2 由大脑活动引导的机器学习

近年来，机器学习和认知神经科学之间的交叉和重叠显著增加。深度学习方法被用于神经响应预测，反过来，诸如编码理论、工作记忆和注意力之类的生物启发机制也越来越多地被采用。然而，到目前为止，人类的认知能力似乎仍然过于复杂，无法通过计算来理解，而数据驱动的“逆向工程”人类思维的方法可能是推进人工智能的最佳方式。

在这种情况下，最近的研究已经使用神经活动数据来限制模型训练。例如，将通过深度前馈模型学习的视觉特征映射到直接从脑电数据学习的大脑特征，以执行自动视觉分类。还有研究使用 fMRI 数据来偏向机器学习算法的输出，并推动其利用在视觉皮层中发现的表征。这项工作类似于第一批依赖大脑活动数据执行视觉分类的方法之一，区别在于前者显式地利用神经活动来权衡训练过程，而后者提出了一种核对齐算法（意思就是

相同大小的卷积核) 来融合视觉分类器的决策和大脑数据。

这两个领域之间其实有更深层的联系: 不是使用神经数据作为信号来权衡计算学习的表征, 而是学习图像和相应神经活动之间的映射, 从而使视觉模式与神经加工一对一地联系起来。这种映射可能会揭示更多关于大脑表征的信息, 并能够以更内在和更全面的方式指导训练加工。因此, 这种方法不仅仅是一种受神经数据启发或约束的混合机器学习方法, 而是一种隐含地发现计算表征、视觉模式和大脑表征之间的相似性, 并使用它们来执行视觉任务的方法。

2.3 多模态学习

现实世界的信息来自多个模态, 每个模态都带有不同形式但同样有用的内容, 用于构建智能系统。具体地说, 多模态学习方法试图通过找到联合表征来学习特征嵌入 (feature embedding), 其本质是通过对不同模态的真实对象的特征学习来寻找能够共同表征的特征单位。有效的联合表征必须保持模态内相似性 (例如, 两个相似图像在联合空间中应该具有接近的矢量化表征; 同样, 两个等价的文本描述也应该具有相似的表征) 和模态间相似性 (例如, 描述该图像的内容的图像和文本片段在联合空间中应该比图像和不相关的文本片段更接近)。根据这一性质, 大多数方法找到视觉数据与文本或音频之间的对应, 以支持区分任务 (例如, 分类) 或以一种模态为条件的对另一种模态的预测 (例如, 图像合成或检索)。对于前一类方法, 已经使用字幕和标签来提高浅分类器和深分类器的精度。类似地, 有的研究使用音频来监督视觉表征; 有的研究利用视觉来监督音频表征; 有的研究使用声音和视觉来共同监督彼此; 还有研究探索了如何通过分析运动和语义线索来分离和定位视频中的多个声音。相反, 其他工作集中于预测来自另一模态的一种模态中的丢失数据, 例如, 从图像生成文本描述, 反之亦然。有研究提出了一种联合表征空间来调节生成对抗网络 (GANs), 用于从文本描述合成图像。有研究使用可变自动编码器从文本字幕合成图像。还有研究使用了一种从大脑信号中学习的嵌入方法, 使用 GAN 和变分自动编码器来合成图像。

在本文中, 使用的方法受到学习共享多模态

表征的方法的启发, 但有几个关键的区别。首先, 使用的一种数据是大脑活动数据 (EEG), 几乎可以肯定它比文本/音频的噪音更大。这使得发现视觉和大脑模态之间的关系变得更加困难。从这个意义上说, 本文方法旨在提高预测的准确性, 并作为一种特征发现工具来揭示大脑加工。因此, 主要目标是学习一个可靠的联合表征, 并探索学习的空间, 找到能够揭示大脑表征的视觉和大脑特征之间的对应关系, 这些特征反过来可以用来建立更好的深度学习模型。

此外, 所提出的由两个编码器 (每个模态一个) 组成的深度多模态网络在孪生网络中被训练, 并且采用损失函数, 该损失函数强制执行所学习的嵌入以表征样本之间的范畴内差异, 而不仅仅是范畴间的区别特征。

3 视觉-大脑特征的多模态学习

神经活动 (由脑电图记录) 和视觉数据有非常不同的结构, 要找到一个共同的表征并不是一件容易的事。以前的方法已经尝试通过训练单个模型来找到这样的表征: 例如, 首先通过训练 EEG 信号的递归分类器来学习大脑表征, 然后训练 CNN 来将视觉特征回归到对应 EEG/图像对的大脑特征。虽然这提供了有用的表征, 但已学习特征的效用与用于计算初始表征的任务紧密相关 (例如, 图像分类), 并且更关注于学习分类特征, 而不是寻找脑电图和视觉模态之间的关系。

因此, 本文认为从人类神经信号和图像到公共空间的任何变换都应该通过最大化每个输入表征的特征嵌入之间的相似性来联合学习。为此定义一种孪生网络, 用于学习使用深度编码器的脑电信号和图像之间的结构化联合嵌入, 并最大化两种模态之间的相似性度量。模型的结构如图 1 所示。设 $\mathcal{D} = \{e_i, v_i\}_{i=1}^N$, \mathcal{D} 是神经信号样本和图像的数据集, 使得每个神经 (EEG) 样本 e_i 响应于观看图像 v_i 而记录在人类对象上。理想情况下, 潜在信息内容应该由 e_i 和 v_i 共享。同样, 设 \mathcal{E} 是 EEG 信号样本的空间, \mathcal{V} 是图像的空间。方法的目的是训练两个编码器, 它们分别将神经响应和图像映射到公共空间 \mathcal{J} , 即 $\varphi: \mathcal{E} \rightarrow \mathcal{J}$ 和 $\theta: \mathcal{V} \rightarrow \mathcal{J}$ 。其思想是通过最大化每个输入表征的两个嵌入之间的兼容性函数来学习公共空间。给定图像和来自一

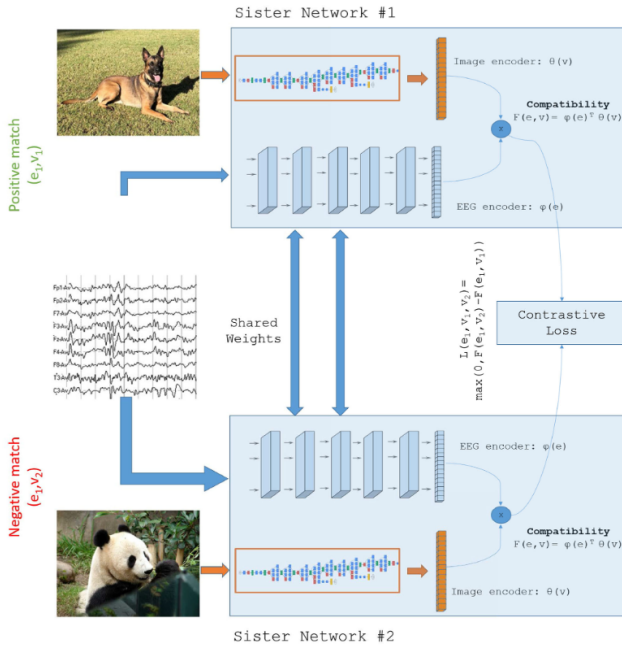


图 1 孪生网络用于学习大脑-图像的联合表征

个对象的相关 EEG 之间的正匹配，以及相同 EEG 和不同图像之间的负匹配，对网络进行训练以确保相关 EEG /图像对之间的相似度比不相关的更接近（更高的兼容性）。

在用于结构化学习的其它方法中，编码器的训练基于兼容性函数 $F: \mathcal{E} \times \mathcal{V} \rightarrow \mathbb{R}$ 的定义的分类问题，该兼容函数计算相似性度量作为 EEG/图像对的各自嵌入之间的点积：

$$F(e, v) = \varphi(e)^T \theta(v)$$

为了从任何特定的任务中提取学习过程，用 triplet loss 来训练孪生网络，目的是将匹配的脑电和图像的特征映射到联合空间中的邻近点，同时排除不匹配的特征。因此，给定两对脑电/图像 (e_1, v_1) 和 (e_2, v_2) ，我们认为 e_1 为锚定项 (anchor)， v_1 为正项， v_2 为负项。使用兼容性 F (其是相似性度量而不是距离度量，相似度可以判断是否正确地学习到了脑电和图像对之间的关系)，用于训练编码器的损失函数变为：

$$L(e_1, v_1, v_2) = \max \{0, F(e_1, v_2) - F(e_1, v_1)\}$$

只有当 (e_1, v_1) 的兼容性大于 (e_1, v_2) 的兼容性时，该公式才指定零损失。请注意，公式中的任何位置都不使用类标签。这确保了最终的嵌入不只是将类别区分向量与 EEG 和图像相关联，而是

试图提取更全面的模式来解释这两种数据模态之间的关系。

3.1 编码器的架构

将神经活动信号映射到联合空间 J 的脑电编码器称为 EEGChannelNet，即在图 2 中给出的卷积网络，具有时间块、空间块和残差块，其遵循分层方法以不同步骤处理输入信号的不同维度。

时间块首先沿着时间维度处理输入信号，对每个通道独立地应用一维卷积，具有提取显著特征和减小输入信号大小的双重目的。时间块的作用是提取表征每个信道内的重要时间模式的信息。

接下来的空间块旨在通过在通道维度上应用一维卷积来寻找相应时间间隔的不同通道之间的相关性。为了阐明这一方面，请注意，大小为 $C \times L$ (其中 C 为通道数， L 为时间长度) 的输入 EEG 信号将由时间块转换为大小为 $F \times C \times LT$ 的张量，其中在应用一维卷积之后， F 为级联特征映射的数目， LT 为“新”时间维度。该张量的每个元素在时间上都不会与原始信号中的单个样本相对应，但会“覆盖”特定的时间感受场，具体取决于卷积核大小。然后，空间块对 LT 维度中每个元素的特征和通道维度进行操作，目的是分析相应时间的空间相关性（在多个尺度上）。与时间块相似，空间块也由多个一维卷积层组成，其输出是级联的。在这种情况下，对通道维度进行排序，以使通道的“行”被连续附加到信号矩阵中；然后，每个空间一维卷积以不同的内核大小运行。时间和空间块中的所有卷积层之后是批处理归一化和 ReLU 激活。一旦模型在时间和空间维度上独立工作，则由一组残差层组成的最终残差块对时空表征进行二维卷积，以从信号中找到更复杂的关系和表征。每个残差层在将输入加到残差之前执行两次卷积（具有批处理归一化和 ReLU 激活）。然后将输出提供给最终的卷积层，然后再提供具有与联合嵌入维数相同大小的完全连接层。

编码器首先经过 EEG 分类测试，方法是在完全连接的编码器层之后适当添加一个 softmax 层，以了解其从神经数据解码视觉信息的能力。然后，使用前面介绍的孪生模式对编码器进行训练。

视觉编码器 $\theta(\cdot)$ 通过卷积神经网络将图像映射到联合空间 J 。使用经过预训练的 CNN 提取视觉特征，并将其馈入线性层以映射到联合嵌入空

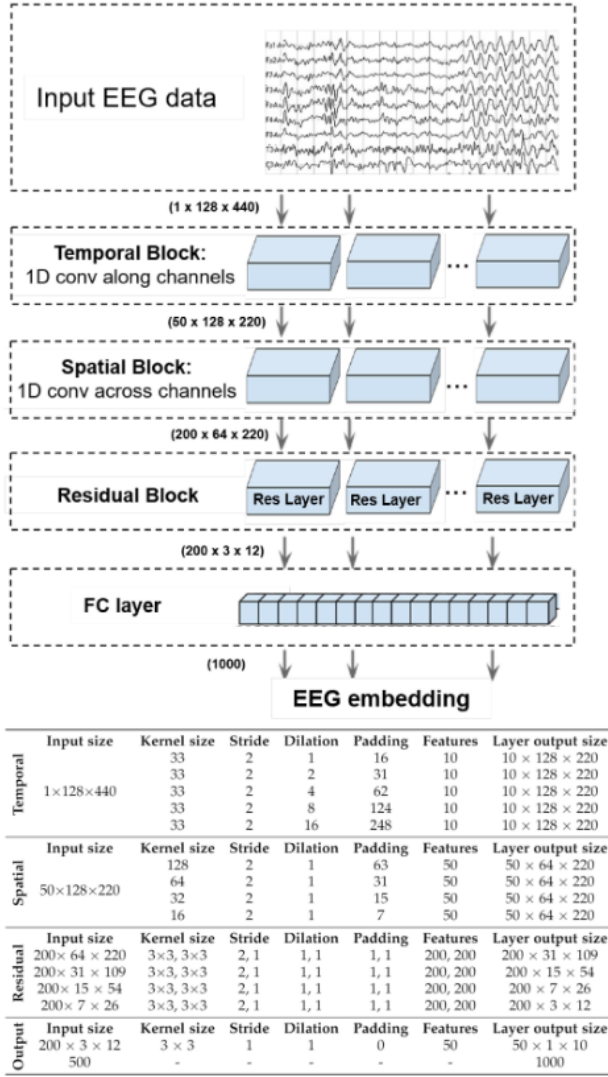


图 2 详细的 EEG-ChannelNet 架构

间。还通过微调图像编码器以端对端的方式学习兼容性功能，以便更好地识别低级和中级视觉大脑表征，并对其进行适当的解码，以提示人们在分析视觉场景时会使用哪些信息。

4 图像分类与视觉显著性检测

孪生网络学习视觉和 EEG 嵌入，以便最大限度地提高图像和相关神经活动之间的相似性。最终可以利用所学的流形来执行视觉任务。在认知神经科学中，有越来越多的证据表明：

- 大脑活动记录包含有关视觉对象分类的信息；
- 注意力影响视觉信息的处理，甚至在灵长类视觉皮质的最早期区域也是如此。特别地，自下而上的感觉信息和自上而下的注意力机制似

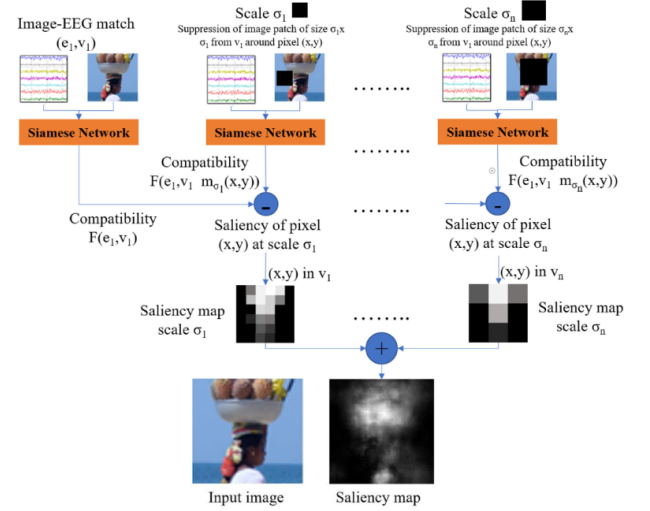


图 3 基于多尺度抑制的显著性检测

乎融合在一个综合的显著性检测过程中，这反过来又分布在视觉皮层中。因此，对视觉刺激作出反应的脑电图记录应同时编码视觉类和显著性信息。

然而，对于图像分类，可以简单地使用训练的编码器作为后续分类层的特征提取器，而对于显著性检测，这里设计了一种基于多尺度抑制的方法，其灵感来自于识别与 CNN 神经元激活相关的像素的方法，该方法分析兼容性度量 $F(1)$ 中的波动。图 3 是显著性检测方法，并且可以形式化如下。设 (e, v) 是与 $F(e, v)$ 相容的 EEG/图像对。像素 (x, y) 和尺度 σ 处的显著性值 $S(x, y, \sigma, e, v)$ 是通过去除 (x, y) 周围的 $\sigma \times \sigma$ 图像区域并计算原始兼容性分数与抑制该补丁后的兼容性分数之间的差来获得的。更正式地说，如果 $m_\sigma(x, y)$ 是一个二进制 mask，其中围绕 (x, y) 的 $\sigma \times \sigma$ 窗口内的所有像素都被设置为零，则有：

$$S(x, y, \sigma, e, v) = F(e, v) - F(e, m_\sigma(x, y) \odot v)$$

\odot 表示按元素相乘 (Hadamard 积)。对于多个比例尺，将像素 (x, y) 的总体显著性值设置为 (每个比例尺) 显著性分数的归一化，然后在逐个图像的基础上执行归一化以用于可视化。

$$S(x, y, e, v) = \sum_{\sigma} S(x, y, \sigma, e, v)$$

5 视觉相关的大脑加工

虽然显著性检测方法可以研究图像中的变化如何反映兼容性分数,但更有趣的是分析神经模式如何作用于学习到的表征。事实上,当掩盖最重要的视觉特征时,可以发现兼容性的巨大差异,类似地可以期望,当从神经活动中删除“重要的”(从视觉特征-匹配的角度出发)成分时,兼容性会下降。进行这种分析传统上需要大脑信号模式的先验知识和人工分析的结合,以这种方式提出问题需要至少部分地了解所观察的加工,这使得自动检测先前未知的信号模式变得复杂。

相反,通过分析兼容性是如何随着输入信号的针对性改变而变化的,这种联合表征可以很容易地将大脑信号与视觉刺激联系起来。因此,与显著性检测相似,可以识别大脑活动中传达视觉信息的空间成分。

人类的目标识别是通过跨皮质区域的形状和特征信息的多级聚合来执行的,从而导致可以很容易地适应接收到的刺激上的各种任务的分布式表征。由于这些原因,了解这种分布的表征如何在人类皮层上进行空间定位是成功模拟人类视觉系统的基本步骤。为了评估每个 EEG 通道(和相应的大脑区域)的重要性,采用学习的联合嵌入空间从 EEG 信号中“过滤”该通道并测量相应的脑电图兼容性图像和滤波后的信号。

每个通道对于单个 EEG/图像对的重要性可以通过计算该对的兼容性分数与从 EEG 信号中抑制该通道时获得的兼容性之间的差值来测量。理想情况下,给定一般的 EEG/图像对 (e, v) , 并用 e_{-c} 表示 e 的变换以便抑制通道 c 上的信息,这里将通道 c 对于 (e, v) 对的重要性定义为:

$$I(e, v, c) = F(e, v) - F(e_{-c}, v)$$

这个方程表达的是抑制传递不必要信息的通道应该会导致兼容性分数的微小差异;类似地,如果一个通道包含重要信息,将大脑活动数据与视觉数据相匹配,那么当该通道被抑制时,兼容性就会下降。

实际上,由于不同替换产生不同的兼容性分数,因此很难找到信道 c 的单个理想替换来计算 e_{-c} 。然而,在大量随机替换的 c 通道上平均兼容性差异会得到稳定的结果:因此,修改了上述方

程,当用随机高斯样本序列替换信道 c 时,计算重要性分数作为兼容性差异的期望值,随机高斯样本序列以 100 Hz 进行低通滤波,并根据原始信道的估计统计(均值和方差)分布。

更正式地说,如果将 EEG 信号 e 表示为每行一个通道的矩阵:

$$e = \begin{pmatrix} e_1 \\ e_2 \\ \dots \\ e_c \\ \dots \\ e_n \end{pmatrix}$$

我们把 $I(e, v, c)$ 计算为:

$$I(e, v, c) = F(e, v) - \mathbb{E} \left[F \left(\begin{bmatrix} e_1 \\ e_2 \\ \dots \\ H(\mathcal{N}(\mu_c, \sigma_c^2)_{L \times 1}) \\ \dots \\ e_n \end{bmatrix}, v \right) \right]$$

其中 μ_c 和 σ_c^2 关注通道 c 的样本均值和方差, L 是脑电时间长度, $\mathcal{N}(\mu, \sigma^2)_{N \times M}$ 是从指定分布中采样的 $N \times M$ 矩阵, H 是 100 Hz 的低通滤波器。

最后,由于在单个 EEG/图像对上计算的信道重要性分数本身对于得出一般结论可能并不重要,因此将信道重要性的定义扩展到多个数据样本上: $I(c) = \mathbb{E}_{(e,v)}[I(e, v, c)]$ 。其中对所有数据集样本计算期望值。

6 解码大脑表征

以前的每一种方法都研究了改变大脑活动信号或图像内容的影响,但它们的局限性在于它们提供的差异分析只在一种模式下进行,本文提供的差异分析可以识别对两个相应编码之间的相似性影响最大的视觉特征,或者可以识别大脑活动中与学习到的表征更相关的空间模式。然而,仍然不知道哪些视觉特征会引起哪些大脑反应,即神经生成器。为了填补这一空白,这里提出了一种额外的解释兼容性差异的方法,通过使用学习的流形来对脑电图通道以及相应的大脑区域进行分析,这些通道是在不同尺度上从边缘到纹理到对象和视觉概念的视觉特征检测中最常用的。为

了进行这一分析, 评估当图像编码器中的特定特征映射被移除时计算的兼容性分数的差异, 并将相应的特征映射到当那些特征被移除时似乎最不活跃(兼容性方面)的 EEG 通道。实际上, 在给定脑电/图像对 (e, v) 的情况下, 将 $F(e, v_{-l,f})$ 定义为通过抑制图像编码器的 1 层的第 f 个特征映射而计算出的兼容函数值。根据前文方程给定的 EEG/图像对 (e, v) , 当去除某一层的特征时, 计算的通道 c 的重要性为:

$$I(e, v_{-l,f}, c) = F(e, v_{-l,f}) - \mathbb{E} \left[F \left[\begin{pmatrix} e_1 \\ e_2 \\ \dots \\ H(\mathcal{N}(\mu_c, \sigma_c^2)_{L \times 1}) \\ \dots \\ e_n \end{pmatrix}, v_{-l,f} \right] \right]$$

然后, 定义成对 (e, v) 的特征 (l, f) 和通道 c 之间的关联如下:

$$A(e, v, c, l, f) = I(e, v_{-l,f}, c) - I(e, v, c)$$

如果在移除给定 (e, v) 对的该通道的固有重要性分数之后, 当该特征被移除时通道 c 的兼容性变化不变, 这将意味着在编码表征中没有留下不匹配的视觉成分, 则认为通道 c 和特征 (l, f) 是相关联的。

可以通过平均通道 c 和层 1 中的所有特征来估计该层之间的关联:

$$A(e, v, c, l) = \mathbb{E}_f[A(e, v, c, l, f)]$$

得出的分数表明在计算模型中的某一层计算的特征与大脑在特定头皮位置加工的特征有多相似。

最后, 对于通道重要性分数, 可以通过对整个数据集进行平均来计算一般关联分数:

$$A(c, l) = \mathbb{E}_{(e,v)}[A(e, v, c, l)].$$

7 结论

在这项工作中, 模型所学习的嵌入使该表征在执行脑活动监督下, 可用于执行多种计算机视觉任务。实验表明, 神经活动可以可靠地驱动图像

分类和显著性检测方法的性能提升。除了推进与脑引导图像分类相关的工作之外, 该方法还提供了一种从 EEG 数据中提取神经表征并将其映射到最重要/最显著的视觉特征的方法。

尽管从这些发现中得出一般的认知神经科学结论不是这项工作的主要目标, 但鉴于认知实验的规模还很小, 文章提出了一种基于 AI 的策略, 该策略似乎可以产生可靠的大脑表征及其相应的头皮活动近似值, 通过共同学习使神经活动与视觉图像之间的相关性最大化的模型。将来这项工作的自然扩展是进一步研究这些关联, 以期发现 EEG 信号与视觉模式之间的更好对应关系, 例如, 通过识别大脑活动中与特定对象, 模式或类别相对应的不同反应, 具有不同的特异性。我认为将人工智能(通过开发更复杂的方法)和神经科学(通过更多量身定制的大规模实验)相结合的联合研究努力是必要的, 通过研究大脑加工如何与人工模型结构相关, 进而利用未发现的神经动力学提出新的神经结构, 使计算模型更接近人类的感知和认知表现。

综上所述, 本文提出的学习大脑加工并将其与视觉线索在时间和空间上相关联的方法产生了两方面的贡献:

1. 人工智能: 引入了新的模型来解码与视觉任务相关的 EEG 信号, 具有最先进的性能, 并以生物学上可信的方式进行解码。此外, 本文的方法允许自动识别与人类神经活动一致的计算特征, 这代表了帮助解释人工智能模型的新方向。
2. 认知神经科学: 本文的方法是朝着为认知神经科学家提供基于 AI 的方法来理解时空神经反应的方法迈出的一步, 而无需设计针对多被试和多试次的实验。当设计出高度精确的人工智能时, 它将允许认知神经科学家模拟人类的反应, 而不是收集大量昂贵的数据。

在学习这门课的过程当中, 随着对认知神经科学和人工智能的深入了解, 我不禁对神经生物学, 脑科学都萌生了兴趣。机器学习中的神经网络其实只是个简单的数学函数, 人类的大脑、细胞的精密程度远超我的想象。希望在有生之年能够见证甚至参与到强 AI 的诞生。